CROATIAN ACADEMY OF ENGINEERING

ANNUAL 2015 OF THE CROATIAN ACADEMY OF ENGINEERING (HATZ)

Published by:
Croatian Academy of Engineering – HATZ
28 Kačić Street, P. O. Box 59, HR-10001 Zagreb, CROATIA

Editor-in-Chief:
Prof. Vladimir Andročec, Ph.D.
President of the Croatian Academy of Engineering

Editorial Board:
Prof. Vladimir Andročec, Ph. D., HATZ President
Prof. Vladimir Medved, Ph. D., HATZ Vice-President
Prof. Zdravko Terze, Ph. D., HATZ Vice-President
Prof. Dubravko Rogale, Ph. D., HATZ Secretary-General
Prof. Emer. Stanko Tonković, Ph. D., HATZ Past-President

Secretary of the Editorial Board:
Melanija Strika, Prof. Soc., HATZ Business Secretary

Assistant:
Andrea Gmajnički, Mag. Iur., HATZ Professional Trainee

Cover Design:
Assoc. Prof. Ivana Žiljak Stanimirović, Ph. D.

ISSN 1332-3482

Annual 2015 of the Croatian Academy of Engineering
Ann. 2015 Croat. Acad. Eng.

Pre-Press:
Vladimir Pavlić, Dipl. Eng. (GRAPA, Ltd., Zagreb)

Printed by:
Tiskara Zelina, Ltd., Zagreb

Circulation:
350 books, 200 CDs

CROATIAN ACADEMY OF ENGINEERING

Ann. 2015 Croat. Acad. Eng. ISSN 1332-3482

ANNUAL 2015 OF THE CROATIAN ACADEMY OF ENGINEERING

Zagreb, 2016

Reviewers

Prof. Vesna Alar, Ph. D.
Prof. Dubravka Bjegović, Ph. D.
Prof. Ružica Čunko, Ph. D.
Prof. Predrag Ćosić, Ph. D.
Prof. Zvonko Dragčević, Ph. D.
Prof. Dr.-Ing. Christoph Gehlen
Prof. Sonja Grgić, Ph. D.
Prof. Marica Ivanković, Ph. D.
Prof. Goran Kniewald, Ph. D.
Prof. Želimir Kurtanjek, Ph. D.
Assoc. Prof. Nenad Leder, Ph. D.
Prof. Sanja Martinez, Ph. D.
Prof. Nikola Mrvac, Ph. D.
Assoc. Prof. Mario Muštra, Ph. D.
Prof. Dr. Peter Neubauer
Assoc. Prof. Eva Ocvirk, Ph. D.
Prof. Klaudio Pap, Ph. D.
Prof. Davor Petrinović, Ph. D.
Robert Precali, Ph. D.
Prof. Dubravko Rogale, Ph. D.
Prof. Marko Rogošić, Ph. D.
Prof. Nikola Rožić, Ph. D.
Prof. emer. Ivo Soljačić, Ph. D.
Prof. Ivica Veža, Ph. D.
Prof. Edita Vujasinović, Ph. D.
Prof. Ivana Žiljak Stanimirović, Ph. D.

TABLE OF CONTENTS

Foreword

Dear Readers,

Croatian Academy of Engineering is a scientific association of the most distinguished and prominent Croatian and international scientists in the fields of technological and biotechnological sciences with the objectives of promoting technological sciences, gathering and encouraging co-operation of the scientists of various technological, biotechnological and other professions in order to support efficient scientific and economic development of Croatia without gaining any profit.

The Academy cooperates with numerous other scientific organizations and institutions in Croatia and abroad as well as with a number of companies in the economy and industry.

As a full member of the Euro-CASE and the CAETS, the Academy has the privilege and the opportunity to transfer the international achievements and experiences in the fields of technological sciences to Croatia as well as to present our achievements to the world.

Therefore one of the essential tasks of the Academy is issuing the publications that summarize and systematize all the aforementioned achievements in the form of books, proceedings, journals, bulletins and similar editions.

Ever since its founding in 1993, the Croatian Academy of Engineering seeks to maintain a continuity of its editions, including the Annuals, the issue of 2015 being presented to you herein.

In the previous year the Academy has published its Jubilee Monograph "Twenty Years of the Croatian Academy of Engineering (HATZ) 1993-2013" in the format of an Annual. The publication provided the most important information on the history of the Academy since its establishment in 1993.

We decided to dedicate the "Annual 2015 of the Croatian Academy of Engineering" to the papers by our members, thus representing their noteworthy scientific and research activities in the previous period.

In the first part of our Annual we present the original papers by our authors, which were prepared for this publication, peer-reviewed and categorized by the distinguished Croatian and international peer-reviewers. The presented fourteen papers are by content parts of a wider spectrum of technological and biotechnological fields, so the reader could, aside from concrete results represented in the papers, gain insight into some of the modern tendencies in certain fields of research, especially those associated with the "state of the art" contemporary technologies in Croatia.

The second part of this Annual represents the re-print of the previously already published notable papers by the members of our Academy. The papers were originally published in journals or conference proceedings in the previous five-year period and the Editorial Board of the Annual accepted them for re-print as a representation of our members' scientific activities. The publishers of the journals and conference proceedings in which the papers were originally published as well as the authors have approved their re-print in our Annual for scholarly usage.

The Editorial Board hereby cordially acknowledges the contributions and efforts of all authors of the papers published in the Annual.

The third part of the present Annual provides a recapitulation of the most important activities of the Academy's bodies and membership in 2015 and gives a very good insight in the scope and engagement of the Academy in pursuing its achievements.

We are particularly pleased to acknowledge herein the cooperation with other distinguished Croatian scientific and cultural institutions: Croatian Academy of Sciences and Arts, Academy of Medical Sciences of Croatia, Academy of Legal Sciences of Croatia, Academy of Forestry Sciences, Miroslav Krleža Institute of Lexicography, University of Zagreb etc.

Prof. *Vladimir Andročec*, Ph.D.
Editor-in-Chief
President of the Croatian Academy of Engineering

Selection of Original Papers by the Members of the Croatian Academy of Engineering

Scientific paper

Supplemented Explanation of the Adequacy of Mathematical Modeling of the Growth Kinetics of Multicellular Aglomerates of Live Organisms Based on the Threedimensional Growth Concept

Marijan Bošnjak[1], Velimir Topolovec[2] and Krešo Mihaljević[3]

[1]Croatian Academy of Engineering, Kačićeva 28, 10000 Zagreb, Croatia
[2]University of Rijeka, Information Science Department, Rijeka, Croatia
[3]Pliva, R&D, TAPI Croatia, Zagreb, Croatia

Relationships referring to the growth kinetics of multicellular agglomerates of live organisms were studied by applying mathematical models suitable for computer simulations. Experimental data from authors previously published papers have been exploited to form more sophisticated mathematical models. Based on previous experience concerning the application of mathematical models, authors described kinetics of supposedly relevant multicellular tumour spheroid growth process events by means of mathematical models suitable for computer simulations. Series of computer simulations was performed to find optimal computer simulation parameter values. Finally, the agreement of experimental with computer simulation data was tested and it was established that computer simulation data fit well to those experimental. Quality of fitting was confirmed after applying statistical validation methods. Results are expressed graphically (Figs.) and numerically (Tables). Applicability of relationships referring to colonies of filamentous microorganisms to other tumour forms is discussed.

Key words:

Multicellular agglomerates, tumour spheroids, microbial colonies, growth kinetics, mathematical modelling, computer simulation.

1. Introduction

In the work submitted for publication under the title „Mathematical Modelling of Probable Process Events Kinetics in Intestinal Compartment of Human Body as a Function of Applied Probiotics“, equations expressing growth kinetics of tumour present in the large intestine have been also included in the developed mathematical model. The equations expressing tumour growth kinetics are based on the already developed mathematical model referring to the growth kinetics of multicellular tumour spheroids MTS (1). This fact asked an additional consideration of mentioned published data. Although in given publication (1) an excellent agreement of already established experimental data with those theoretical based on the three-dimensional growth kinetics concept, previously proven as convenient for describing growth kinetics of some filamentous microbes (2,3), the already published data (1) suggested an additional their exploitation for more complete explanation of growth kinetics relationships. To realize an advanced approach to the developing of more sophisticated mathematical model, it appeared necessary to exploit again the experimental data present in already published paper (1). However, these are presented graphically and their reinterpretation becomes possible after an estimation of their corresponding numerical values.

Fortunately, since the values of corresponding kinetics constants in the same paper (1) are expressed numerically, the danger of possible inadequate estimations of numerical values of graphically expressed experimental data can be neglected. Prior to starting the description of extended mathematical model one should point out the fact that studies of the growth kinetics of multicellular tumour spheroids followed as the consequence of difficulties connected with mathematical descriptions of enormously more complex characteristics of the growth kinetics of tumours developing in live organisms (4,5). This means that reasons for developing mathematical model based on the three-dimensional growth concept should be mentioned again. In the published text (1) it was pointed out that an *in vitro* method of multicellular tumour spheroid (MTS) culture was developed to overcome certain limitations of monolayer cultures. Consequently, the information followed that within limited range of cultivation conditions the radial growth of MTS could be well expressed as a linear function of cultivation time (4-7). However, the Gompertzian equation expressing growth kinetics of the *in vivo* tumour cell growth, and fitting the experimental data better than the exponential growth model, appeared (1,8). Experimental points, however, deviated markedly from the calculated curves. As mentioned, the mathematical model we developed (1) was based on the three-dimensional growth kinetics concept which was previously verified when describing growth kinetics of microbial cells belonging to some genera of bacteria (2, 3). Later, confirmation was extended (***e.g.*** to Streptomycetes and some species of *Aspergillus* genus (9-12)).

Repeated exploitation of experimental data applied in the previously published paper (1) and actually desired further improvement of already developed mathematical model (1), ask a corresponding repeated description of particular steps of applied procedure of mathematical modelling. Therefore, one can start with defining the growth kinetics during the cultivation conditions when growth kinetics restriction does not play remarkable role, *i.e.* when three-dimensional growth kinetics can be expressed by differential equation:

$$\mathrm{d}X/\mathrm{d}t = k_1 \cdot X^{2/3} \tag{1}$$

In cell cultures in general, the growth kinetics is a function of the number of cells, biological and physical characteristics of cells, properties of culture medium and its close environment, and cultivation conditions as well. During their growth cells consume nutrients from the medium and produce metabolites which interfere with growth of cells. Total volume of particular cell agglomerates formed as multicellular spheroids is proportional to the third power of their diameters, whereas their total surface area is proportional to the second power of their diameters. Contacts with nutrients are achieved on spheroid surfaces, and therefore the growth rate of spheroids under unrestricted growth conditions can be expressed by equation (1). However, since during the growth in the closed reaction systems MTS become exposed to more and more unfavourable growth conditions, their specific growth rates continuously reduce. Therefore, for such conditions one should apply the equation

$$\mathrm{d}X/\mathrm{d}t = k_1 \cdot X^{2/3} - k_2 \cdot X \tag{2}$$

Its applicability was proven in describing kinetics of mycelia forming micro-organisms already during the second half of 20th century (7-10). As a consequence of continuous reduction of growth rate the equilibrium state of rates will be established after the maximal agglomerate volume becomes attained. Then the following condition becomes fulfilled

$$\mathrm{d}X/\mathrm{d}t = k_1 \cdot X_m^{2/3} - k_2 \cdot X_m = 0 \tag{3}$$

and, therefore also

$$k_1 \cdot X_m^{2/3} = k_2 \cdot Xm \tag{4}$$

Since consequently it follows that relation between kinetic constants becomes determined by equation

$$k_1 = k_2 \cdot X_m^{1/3} \tag{5}$$

the rate of agglomerate volume changes can be defined by equation

$$dX/dt = k_2 \cdot (X_m^{1/3} \cdot X^{2/3} - X) \tag{6}$$

Upon the separation of variables an integral form of equation can be obtained, *i.e.*

$$\int dX/(X_m^{1/3} \cdot X^{2/3} - X) = \int k_2 \cdot dt \tag{7}$$

As demonstrated originally (7) and repeatedly (1), the integration in the interval (X_0, $-X$) and (t_0, $-t$) respectively, gives:

$$3/(t - t_0) \cdot \ln[(X_m^{1/3} - X_0^{1/3})/(X_m^{1/3} - X^{1/3})] = k_2 \tag{8}$$

or

$$6.9/(t - t_0) \cdot \log[(X_m^{1/3} - X_0^{1/3})/(X_m^{1/3} - X^{1/3})] = k_2 \tag{9}$$

Application of equation (9) for expressing growth kinetics of MTS asked equation transformation. Since the volume of MTS is proportional to the third power of its diameter, *i.e.*

$$X = k \cdot D^3 \tag{10}$$

the following equation can be applied:

$$6.9/(t - t_0) \cdot \log[(D_m - D_0)/(D_m - D)] = k_2 \tag{11}$$

The equation (11) was used for the evaluation of cell growth through diameter values (1). Based on experimental values of MTS diameters several values for maximal diameter D_m were assumed and the corresponding values of k_2 were calculated applying equation (11). It was found that the D_m of 1.6 (mm) and the calculated k_2 of 0.363 (day – 1) fit well to experimental results. Then, the agreement of experimental MTS diameter values (ordinate, y-axis) with those calculated (abscissa, x-axis) was validated applying the statistics of linear regression analysis. Values of straight line coefficient (a = 0.98923), y-axis intercept (b = 0.01128) and of correlation coefficient (r = 0.99696) confirmed an excellent agreement of theoretical with experimental data. The findings were demonstrated also graphically.

The volumetric MTS growth kinetics was simulated according to equation (6). The already calculated k_2 value of 0.363 (day – 1) was used according to relation (5) to calculate k_1 value, and the value of 0.468 resulted. For calculating X_m and X values the expressions

$$X_m = \pi/6 \cdot D_m^3, \text{ and } X = \pi/6 \cdot D^3 \tag{12}$$

have been applied. Experimental (X_E) values were calculated on the basis of experimental MTS diameter values. Simulated and experimental values were compared and the quality of fitting investigated. Statistics of linear regression analysis was applied. The excellent agreement between theoretical and experimental data was established ($a_v = 0.97927$, $b_v = 0.01108$ and $r = 0.99012$).

A similar procedure was applied for evaluating the growth kinetics on the number of viable cells basis. Although the experimental estimation method of numbers of viable cells (cells capable to survive) within MTS of different age was accompanied with difficulties, a good agreement between theory and experiment was observed when applying computer simulation ($k_{1N} = 105.7$ and $k_{2N} = 2.7$) for an investigation of the dependence of viable cell numbers on MTS age.

The study asked an insight into relationship expressing how specific volumetric cell numbers (viable cell concentration) depend on reciprocal MTS diameters. It was established (1) that the numbers of viable cells per unit of MTS volume were decreasing with the increase of MTS diameters. For interpolated linear relationship expressing the dependence of specific viable cell numbers on reciprocal MTS diameters the correlation coefficient of 0.94564 resulted. Extrapolated intercept value on x-axis was found to correspond to MTS diameter of 1.57 mm, which was close to the previously estimated Dm value of 1.6 mm.

The mentioned relationships and established findings (1) can serve to form an integral mathematical model describing the kinetics of studied events during MTS growth.

Mathematical model

Growth kinetics of viable MTS cells:

$$dX_v/dt = k_{x1} \cdot X_v^{2/3} - k_{x2} \cdot X_v - k_{x3} \cdot X_v \tag{13}$$

$$dX_{nv}/dt = k_{x3} \cdot X_v - k_{x4} \cdot X_{nv} \tag{14}$$

Kinetics of other events:

$$dXt/dt = dX_v/dt + dX_{nv}/dt \quad (15)$$

$$dN/dt = k_{n1} \cdot N^{2/3} - k_{n2} \cdot N - k_{n3} \cdot X_{nv} \quad (16)$$

$$dD_v/dt = k_{Dv1} - k_{Dv2} \cdot D_v \quad (17)$$

$$dD_{nv}/dt = k_{Dnv1} - k_{Dnv2} \cdot D_{nv} \quad (18)$$

$$dR/dt = -k_{r1} \cdot R \quad (19)$$

To compare how MTS viable cells relate to corresponding MTS volumes, the additional differential equation is included in the mathematical model:

$$dR_1/dt = -k_{r2} \cdot R_1 \quad [20]$$

Computer simulation

In accordance to previous successful applications (9, 12-14) of Scientist computer simulation program (Micromath, St. Louis, MO, USA) it was also applied in this work. On the basis of mathematical model developed in this work, adequate computer simulation kinetic model was prepared and applied. Fittings of computer simulation to experimental data were statistically validated applying the Jacobian matrix installed as part of Scientist calculation program.

Results and Discussion

Results of computer simulation are demonstrated in Figs.1-3, whereas data referring to statistical validation of the fitting of simulated to experimental data are disclosed in Table1 and partly in Fig.3 and Table2.

An insight into Fig.1 and Fig.2 evidently suggests the agreement of computer simulation with experimental data is satisfactory, and that the participation of nonviable cells increases with MTS age, whereas the data of statistical validation of quality of this agreement undoubtedly confirmed the adequacy of applied mathematical model. Fig.2 differs from Fig.1 by added curve representing kinetics of changes of R1 values, *i.e.* of theoretical (simulated) relative viable cells concentration with

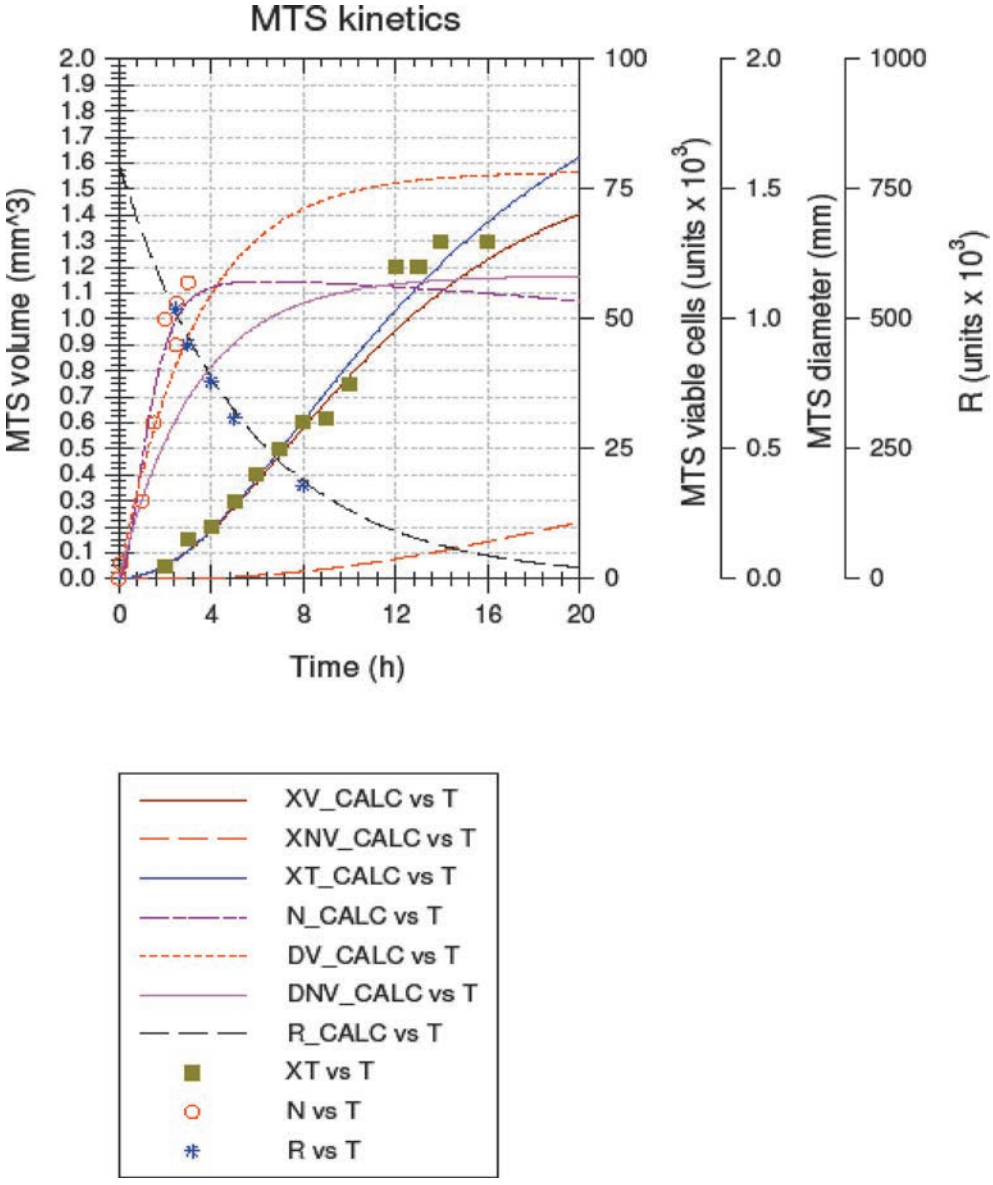


Fig. 1 – Growth kinetics of multicellular tumor spheroids as function of time. Applied simulation kinetics constant values: k_{x1} = 0.5; k_{x2} = 0.405; k_{x3} = 0.017; k_{x4} = 0.020; k_{n1} = 137.0; k_{n2} = 3.55; k_{n3} = 23000.0; k_{Dv1} = 0.47; k_{Dv2} = 0.3; k_{Dnv1} = 0.35; k_{Dnv2} = 0.30; k_{r1} = 0.18

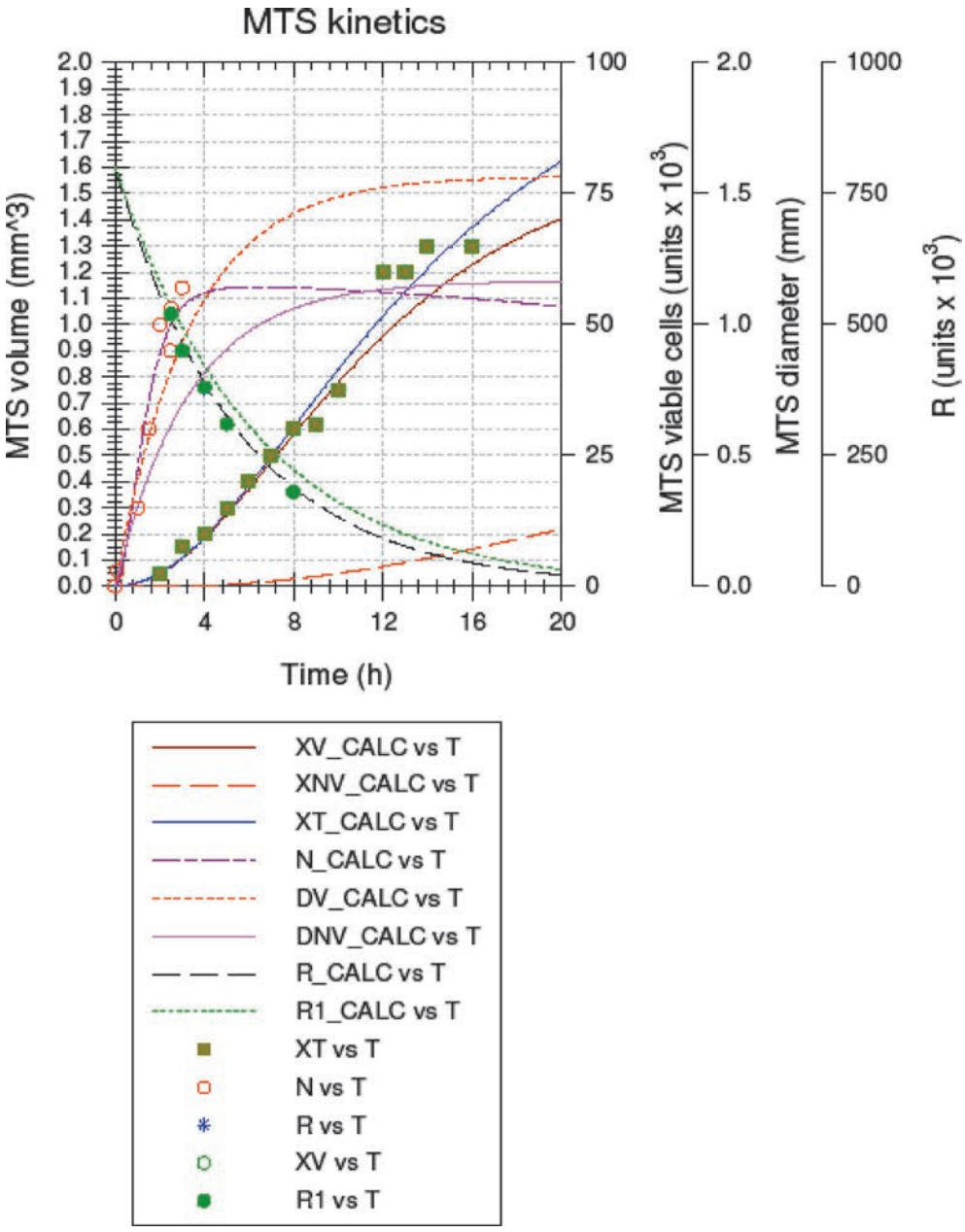


Fig. 2 – Growth kinetics of multicellular tumor spheroids as function of time. Applied simulation kinetic constant values: k_{x1} = 0.5; k_{x2} = 0.405; k_{x3} = 0.017; k_{x4} = 0.020; k_{n1} = 137.0; k_{n2} = 3.55; k_{n3} = 23000.0; k_{r1} = 0.18; k_{r2} = 0.16; k_{Dv1} = 0.47; k_{Dv2} = 0.3; k_{Dnv1} = 0.35; k_{Dnv2} = 0.30

Table 1 – Fitting computer simulation to experimental data validated by applying Jacobian matrix calculation program, installed as subprogram of Scientist program

Data referring to variable	Correlation coefficient	Determination coefficient	Model selection criterion	Fig.
Total MTS volume	0.987436393	0.973754678		Fig.1
MTS viable cells	0.977371298	0.954271620		
Relative viable cells concentration	0.997276227	0.990100458		
Data set	0.999373021	0.998657921	5.72464686	
Total MTS volume	0.987436393	0.973754678		Fig.2
Viable MTS volume	0.986815926	0.942768611		
MTS viable cells	0.977371298	0.954271620		
Relative viable cells concentration (R)	0.997276227	0.990100458		
Relative viable cells concentration (R1)	0.996343475	0.902734297		
Data set	0.998100416	0.993632102	4.49126843	

reference to viable MTS volume. These values evidently differ more from experimental values estimated for total MTS volumes, which have been used to calculate statistical values referring to fitting of computer simulation to experimental data of both R and R1 parameters. As expected, data disclosed in Table 1 clearly show that the same experimental data fit better to theoretical R than R1 values, suggesting that really nonviable cells participate in total MTS volumes. Data in Table 1 in general, confirm the adequacy of applied mathematical model. Since the values of determination coefficient express how much changes of experimental values could be explained by corresponding changes of theoretical values, one can conclude the applied mathematical model appeared as quite adequate. When data set for total MTS volumes is considered, such a conclusion appeared as quite acceptable. These findings confirmed those previously established (1) and with resulting new information concerning the relative distribution of viable and nonviable cell masses in total MTS biomass.

The agreement of simulation with experimental data can be validated also by statistics of linear regression analysis, as demonstrated recently in the paper describing the kinetics of microbial oxidation process events with reference to L-sorbose formation in a large range of culture conditions (14). In the paper, normalized experimental data were compared with corresponding theoretical. In present work, such an approach to the comparison of experimental with computer simulation data led to the Fig.3. As clearly demonstrated in Fig.3, evidently there is a quite accept-

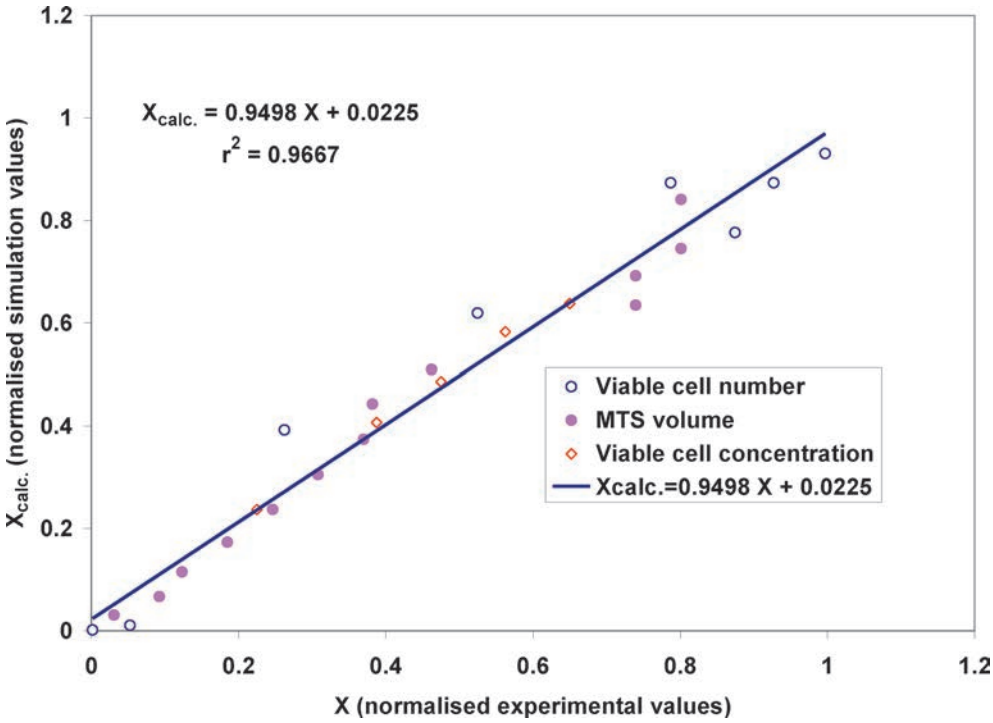


Fig. 3 – Fitting of MTS kinetics normalized experimental to MTS kinetics normalized computer simulation data

able agreement between all experimentally established and their corresponding theoretical (simulated) data. Very low ordinate intercept value (0.0225) and the straight-line coefficient of 0.9498 (very close to 1.0) testify statistically close to that ideal functional dependence between two variables. Moreover, determination coefficient (r^2) value of 0.9667 confirms such a conclusion, since 97 percent of changes of experimental data can be explained by corresponding computer simulation values. More detailed insight into agreement between experimental and theoretical data is given in Table 2. Numerically the highest value of determination coefficient refers to the MTS viable cell concentration, quite as expected after an insight onto curves referring to R values presented in Figs. 1 and 2. No doubt, the satisfactory agreement between compared data is confirmed.

Based on established good agreement of experimental with theoretical data it could be useful to extend the discussion towards more sophisticated explanation of events connected with growth of MTS spheroids in order to facilitate explanations of kinetics of tumours growing in live organisms and therefore in human bodies.

Let consider the growth of MTS is the consequence of genetically predetermined action of the *vis vitalis* of tumour cells against their substrate where they grow. This

Table 2 – Agreement of mathematical model and normalized experimental data (Equation of linear regression: $X_{calc.} = a \cdot X_E + b$)

Quantities	Simulation value range	Straightline slope a	Ordinate intercept b	Determination coefficient r^2
Viable cell number	0.0017491 – 0.997009	0.9294	0.0454	0.9553
MTS volume	0.0308166 – 0.8012327	0.9563	0.0088	0.9751
Viable cell number / MTS volume	0.225 – 0.650	0.9430	0.0343	0.9946

means that we can imagine MTS as expanding bioreactor. Contacts between MTS and its substrate realize over MTS outer cell layer surface, whereas the internal force of expanding MTS acts on internal surface of MTS viable cell layer. With increasing the total MTS size the external and internal cell layer surfaces of outer MTS layer increase as well. However, relative viable cell layer deepness becomes reduced with MTS age increase. With increasing the participation of nonviable cells the MTS autolysis potential is also becoming increased (As already mentioned the long-time known (15), enzymes playing a catalytic role in cell reproduction act autolytically on cells after establishing inconvenient conditions for cell growth and reproduction). Since viable cell layer resistance to its interruptions at outer MTS surface decreases with MTS growth, due to internal pressure onto internal and external surfaces of the outer MTS cell layer, and depends on the deepness of the same, outer cell layer interruptions at its external surface probably will result after critical MTS age is reached. After outer cell layer interruptions at outer MTS surface the formation of new MTS becomes possible, if existing substrate is convenient for cell growth. It is important to point out that mean cell age in MTS layers decreases from centre to periphery. To some extent analogous kinetics relationships could be supposed concerning some tumours developing in human body tissues. However, the real situations referring to tumours growing in human body tissues are much more complex, even when concerning only to tumour size and forms.

It is important also to point out that the role of viruses in the appearance and growth of tumours in live organisms should not be neglected. Data which can be obtained by studies of culture systems of microbial cells could also be helpful in facilitating studies of enigmas of such a kind. In the recent paper referring to the study of process kinetics of bacterial and mould cells infected by phages or fungal viruses respectively, useful information can be found (13). Therefore, one could recommend the reading of the whole content of mentioned paper.

Remarks that characteristics and growth kinetics of tumour cells and microorganisms should not be compared are possible. However, since authors of this work

established the growth kinetics relationships valid for cell agglomerates of mycelia microorganisms can also be applied to describe adequately growth kinetics of MTS, there is reason to suggest the further comparison trials, because one of advantages of applied mathematical model is in its convenience to be transformed into mathematical model describing growth kinetics of multicellular tumour cell agglomerates of other forms, like those of colonies of streptomycetes. An excellent example for exploitation towards such trials seems to be the paper quoted here (9), especially because in it the growth kinetics during lag phase is adequately described. Commonly, one can consider the starting tumour growth phase is its lag phase. By the way, one should mention that the other known mathematical models describing tumour growth kinetics probably could also be transformed into those convenient to describe growth kinetics of tumours of different forms.

To facilitate the understanding of the complexity of events connected with tumours present in human body, different approaches to describing their growth kinetics should be recommended. Some of them (already known) can be mentioned in this discussion: a) tumour developing in human body grows in tissue of normal cells which serves to it as a carrier and nutrition substrate; b) tumour cells and normal tissue cells grow simultaneously competing for substrates and producing metabolites acting differently on cells of different properties; c) normal tissue cells and tumour cells both ask energy for their growth and viability maintenance; d) cells for their reproduction ask an adequate availability of phosphates which can be used for building nucleic acids; e) normal tissue cells and tumour cells are both supplied with nutrients by organism cardiovascular system. It enables a transfer and an elimination of metabolites produced by cells, and plays the fundamental role when application of therapeutic substances becomes necessary to inhibit unwanted biochemical processes;

According to the known, probable frequency of the appearance of tumours in human population (the same is supposed to be considered valid for mammalians generally?) increases with organism body physiological age, especially after mature age is reached. This could mean that, to some extent, tumour growth kinetics could be controlled by reducing energy metabolism to a level of normal tissue cells maintenance; g) If a said could be accepted, perhaps the approaches analogous to those applied in therapies of bacterial infections could be recommended to be applied for body therapy against tumour growth. It is known that, *e.g.* beta-lactam antibiotics act mainly against cell wall formation of sensitive bacteria; h) No doubt, advances in the most important directions of cancer research are evident. Therefore the probability for more successful continuously increasing new scientific findings and discoveries of new drugs could be expected; i) Common disadvantages of drugs result due to their inadequate selectivity. To reduce this problem perhaps a more efficient screening of more adequate drugs could be realized by applying mechanically fortified solid media for tissue cultures, designed analogously to those applied for

screening microorganisms (13), and adapted for normal and tumour cell cultures. Advantage of mechanically fortified solidified media results due to their convenience for using both media layer sides: one for active substance sources installation (cultivation) and the other (reversed) for investigated cell cultures; j) Obstacles weighting successful treatment of tumours should be recognized. A. Marusyk and K. Polyak recently established that intra- tumour heterogeneity is a major obstacle in successfully eradicating tumours (17). From the kinetic viewpoint this could mean that populations of tumour cells should be considered as systems of mixed cell cultures; k) Series of other properties of tumour cells should also be taken into account when approaching to mathematical modelling of tumour growth kinetics in organisms of mammalians. When concentrating to events in the one of organism body compartments, the relevance of event kinetics in other compartments should be taken into account in the extent of probable their influence. The most recent information referring to folate receptor structure is encouraging. One considers that structure, prominent in cancer cells, could lead to targeted therapies (18). Therefore, one can expect a more relevance to be given to studies of the kinetics of drug transfers to corresponding receptors. Although one cannot expect the mathematical modelling would solve problems referring to therapy of tumours, there is no doubt that adequate mathematical models could facilitate a solving of such problems. Therefore, perhaps it could be useful to mention that based on the already long-time known, and postulated later (19), one can approach to mathematical modelling taking into account roughly expressed the following (19,20): “Our cognitions are the reflections of reality, and represent a simplified image of the part of reality. Mathematical model as a reflection of our cognitions on the part of reality can also be the simplified image of the part of reality expressed mathematically”. From the viewpoint of system analysis, complex reaction systems can be represented by their interconnected subsystems. This should be valid also for complex systems of different *in vivo* developing tumour cell agglomerates, ***i.e.*** tumours developing in organisms.

The results of this work could be considered as encouraging the trials of describing tumour growth kinetics by mathematical models appropriate for computer simulations. Recent announce of C&EN news of the week (October 14, 2013) refers to 2013 Nobel Prize in Chemistry Awards: “Karplus, Levitt and Warshel honoured for modelling complex chemical systems and their approach for simulating processes in bio-molecular systems”. Therefore, there is reason to expect advances resulting due to trials of computer simulation even when testing different hypotheses. Commonly after preliminary trials the more successful trials should follow. However, more of optimism concerning successful therapies perhaps could appear after reading of the just now published book of N. Zurak (21). Concerning the possible beneficial effects of microbial probiotics with reference to tumour therapy efficiency, perhaps it could be fruitful to study process kinetics of mixed cultures of multicellular tumour spheroids and microbial cells expressing probiotics activities.

References

[1] M. Bošnjak, V. Topolovec, J. Sorić, L. Milas, Mathematical modelling of growth kinetics of multicellular tumour spheroids, in: P. Gugić, , D. Orlić, (eds.), *Proceedings of the Medicine and Technique Symposium*, *JUREMA* ***24*** (1979) 51-55, Zagreb, 1979.

[2] M. Bošnjak and V. Johanides, Kinetika rasta *Streptomyces rimosus* T6 i biosinteza oksitetraciklina (in Croatian), *Mikrobiologija* (YU), ***10*** (1973) 179-188

[3] M. Bošnjak, V.Topolovec and M. Vrana, Growth kinetics of *Streptomyces erythreus* during erythromycin biosynthesis, ***J. appl. Chem. Biotechnol. 28*** (1978) 791-798

[4] J. M. Yuhas, A. P. Li, A. O. Martinez, and A. J. Ladman, A simplified method for the production and growth of multicellular tumor spheroids, *Cancer Res.,****37*** (1977) 3639-3643

[5] J. M. Yuhas, and A. P. Li, Growth fraction as the major determinant of multicellular tumor spheroid growth rates, *Cancer Res., **38*** (1978) 1528-1532

[6] J. M. Yuhas, A. E. Tarleton and K. B. Motzen, Multicellular tumor spheroid formation by breast cancer cells isolated from different sites. *Cancer Res.,****38*** (1978) 2486-2491

[7] J. M. Yuhas, A. E. Tarleton and J. G. Harman, *In vitro* analysis of the response of multicellular tumor spheroids exposed to chemotherapeutic agents *in vitro* or *in vivo*. *Cancer Res., **38*** (1978) 3595-3598

[8] Z. P. Pavelić, G. W. Porter, L. M. Allen and E. Mihich, Cell population kinetics of fast- and slow- growing transplantable tumours derived from spontaneous mammary tumours of the DBA/2 Ha-DD Mouse. *Cancer Res. **38*** (1978) 1533-1538

[9] M. Bošnjak, A. Bago Joksović, J. Pigac, Ž. Bošnjak Cihlar, D. Hranueli, Applicability of mathematical models in defining the behaviour kinetics distinction among microbial strains, Chem. Biochem. Eng. Q. ***20*** (2006) 375-388.

[10] R. Valinger, J. Beljak, M. Bošnjak, M. Ćurčić, Lj. Vitale (1981): Kinetics of virus replication and glucoamylase biosynthesis in *Aspergillus niger* repeated fed-batch culture, Adv. Biotechnology ***1*** (1981) 229-234

[11] M. Bošnjak, J. Pigac, R. Valinger, M. Vampola, M. Vešligaj, Kinetics of product formation in microbial colonies, in: A. Blažej, A. Ottovaa (eds.), *Progress in Biotechnology Vol. 6* (Proc. Interbiotech ’89), Elsevier, Amsterdam, 1990, pp.345-358

[12] M. Bošnjak: Uvod u kinetiku mikrobnih procesa (Introduction to Kinetics of Microbial Processes (in Croatian)), Graphis, Zagreb, Croatia (2009).

[13] M. Bošnjak, D. Hranueli, J. Maertens, D. Valinger, Ž. Kurtanjek, E. J. Vandamme: Mathematical modelling of virus replication kinetics as a function of the biological reaction system properties, in: Z. Kniewald (Ed.) Annual 2013 of the Croatian Academy of Engineering

[14] K. Mihaljević, M. Bošnjak, Defining of the kinetics of microbial oxidation process events with reference to L-sorbose formation in a large range of culture conditions, Chem. Biochem. Eng. Q. ***23*** (2009) 239-250.

[15] I. Groš, M. Bošnjak, V. Johanides, V. Topolovec, Autolysis of *Streptomyces erythreus,* First Eur.

[16] Congr. Biotechnol. (Interlaken, Switz., Sept. 25-29,1978), Preprints, Part1, Discussion papers, 3/69, DECHEMA, Frankfurt/M, 1978

[17] M. Bošnjak, Ž. Bošnjak Cihlar, D. Kirša, Advantages and perspective of fortified agarose media application, Chem. Biochem. Eng. Q. ***20*** (2006) 217-223.

[18] A. Marusyk, K. Polyak, Cancer cell phenotypes, in fifty shades of grey, Science 339 (2013) 528-529, www.sciencemag.org

[19] CEN.ACS.ORG, 5, July 22, 2013 (Ed. William G. Schulz and Craig Bettenhausen)

[20] N. W. F. Kossen, Mathematical Modeling of Fermentation Processes: Scope and Limitations, in A. T. Bull, D. C. Elwood, C. Ratledge (Eds.), Microbial Technology: Current State, Future Prospects, 29th Symp. Soc. Gen. Microb., Cambridge, University Press, Cambridge, 1979, pp. 327-357.

[21] M. Bošnjak, Matematičko modeliranje biokemijskih reakcijskih sustava (Mathematical modeling of biochemical reaction systems (in Croatian)), Kem. Ind. ***31*** (1982) 545-559

[22] N. Zurak, Arhetip granice (in Croatian, available), Border Archetype (in press), Medicinska naklada, Zagreb, Croatia (2013).

Scientific paper

Impact of Culverts on Sea Water Exchange in Marinas

Aleksandra Babić[1,*], Vladimir Andročec[2,*], Goran Lončar[3,*], Dalibor Carević[3,*]

[1]PROTOK d.o.o.
[2]Akademija tehničkih znanosti Hrvatske
[3]Građevinski fakultet sveučilišta u Zagrebu

This study demonstrates the results of 3D numerical simulations carried out on seawater circulation and exchange with the aim of quantifying the contribution of pipe culverts and wind in the exchange of 'old' seawater from a marine aquatorium with 'exterior' seawater. As a comparative parameter, so-called 'e-folding' time was used for determining the quality of individual pipe culvert solutions. The applied research methodology defined the optimal position for (implementing) culvert construction in characteristic environmental conditions (intensity of tidal signal and wind activity from different directions and with different strengths) as well as the detection of areas with longer-term seawater retention in the marine aquatorium. Within the implementation of the numerical simulation besides the culvert position, basic geometric marine features (input size, protruding or regressed position of marina) were varied.

The results demonstrate that the most intensive seawater exchange is stimulated by the construction of culverts in the breakwater opposite the marina entrance. Wind activity contributes significantly to seawater exchange; in particular, winds from the directions SE and NE ensure e-folding time reductions for by an average of 26% (SE) and 28% (NE) in relation to conditions without wind action. An increase of marina entrance by 100% improves the seawater exchange to a lesser extent then less than culverts implementation.

Key words:
marinas, numerical model, seawater exchange

[*]aleksandra.babic@protok.com, androcec@hatz.hr, goran.loncar@grad.hr, car@grad.hr

1. Introduction

Nautical tourism is defined from a legal point of view as touristic navigation and accomodation on vessels, as well as accomodation in nautical tourism harbours for the purposes of leisure and recreation. Because of its indented coastline and large number of islands Croatia is exceptionally suited to the development of this touristic component. Due to these many islands, the Croatian coastline is protected from open-sea large waves which, in turn, makes possible the building of small-scale harbours. The development of nautical tourism demands the provision of the following coastal services: protection for boats from storms weather, safe harbour and facilities wherein the primary role of marina, port and harbour is the protection of vessels from unwanted sea action caused by waves. The range of capability of marinas/ports/harbours in the broader context is realized in the implementation of maritime construction (protective structures) with which wave intensity in the protected aquatorium may be lessened. The most important criterion for consideration in the selection of a site for a future marina is the location, its natural degree of protection and water depth. Bearing in mind that the breakwater is the most significant part of the whole marina, the aim is to reduce its size to as much as possible by selecting a partially or wholly protected location. On the other hand, as the whole breakwater *size* grows progressively with sea depth, it is therefore economical to consider a location with a water depth of less than 10 m. The relatively mild wave climate of the Croatian Adriatic coastline allows for the construction of breakwaters of smaller dimensions (height and width) than those that are customary on exposed coastlines of the Mediterranean or open oceans. Among the most often used types of breakwater found in Croatia are the bulk (rubble mound) breakwater, the wall-type breakwater, the pilot breakwater with AB wave barrier, the semipermeable gravitational breakwater with AB wave barrier and the pontoon breakwater.

Regardless of the design method of protective structures (pontoon, partially submerged wave barrier, breakwater, wave breaker etc.) the desired reduction of wave intensity is achieved by damming of the sea in the profile of final length and depth. By such action current intensity is reduced in the enclosed body of water and consequently there is less damage to the surrounding natural environment. In order to avoid or reduce unwanted consequent building demands on the main ecosystem of the aquatorium within the ‘new’ project it is anticipated that, for example, the breakwater implementation in the form of pilot construction with partially immersed screens to appropriate depth which is needed to design the breakwater (examples in Umag and Dubrovnik – Gruž). Such solutions in principle are acceptable in mild wave climate conditions and great depth in which it is necessary to implement breakwaters. On the other hand, Croatia has the greatest number of marinas, harbours and ports protected by bulk or gravitational breakwaters which physically

separate the marine aquatorium from the surrounding seawater and prevent natural circulation and water exchange. There is increasing demand on marina water quality due to the fact that boat users regularly stay in their own craft in marinas. Recognition of the problem of reduced seawater circulation in closed aquatoria is the result of the introduction of practice in planning and constructing culverts through the body of bulk and gravitational breakwaters. The purpose of constructing culverts is the forcing of seawater exchange between the enclosed aquatorium and the 'open' sea. However, according to our findings, to date they have not been conducted on the basis of research into the contribution of pipe culverts to seawater exchange. This paper demonstrates the results of numerical simulations of seawater circulation and tracer solution transport, carried out in order to quantify the contribution of culverts to the exchange of 'old' seawater in marine aquatoria with 'outside' water.

As a comparative parameter, so-called 'e-folding time' was used for determining the quality of individual culvert solutions. For defining 'e-folding' time a detailed approach is used, described in the work of Cucco and Umgiesser (2006). Initially, a dimensionless concentration of tracer (nonreactive) is placed in the whole area of the enclosed harbour basin, and the concentration value 0 for the remaining part of the model spatial domain. Following seawater exchange there is a dilution of the initial concentration through the mechanism of convective dispersion, in relation to the fall in average values of the tracer concentration in the harbour basin area. With the adoption of such a method, detection of 'dead zone' long-term retention areas (that is, areas of increasing tracer concentration) is made possible. By the application of numerical simulations, time series of averaged tracer concentrations in the enclosed aquatoria were obtained. By analysing the results of each simulation carried out, the associated 'e-folding' times were obtained, which represent the time necessary for tracer concentration to reach 37% of its starting value. The results obtained in this study may serve designers in making decisions about the needs of culvert construction.

2. Materials and methods

For the implementation of the numerical simulation the 3D numerical model Mike 3 (www.dhigroup.com) was used. The spatial domain (figure 1) was discretised with resolution $\Delta x = \Delta y = 2$ m in the horizontal and $\Delta z = 1$ m in the vertical direction. The bottom has a horizontal constant depth of 5 m. Culverts were defined as openings with width of 4 m and length of 2 m. The culvert lower elevation was positioned at a depth of 0.5 m from the mean sea level. In this way at total $2m^2$ inflow surface for each culvert (at mean sea level) was obtained. Along the break-

water, independently of its configuration and position, a total of 5 culverts were placed (total 10 m² opening at mean sea level).

In terms of marine geometry variation and environmental conditions the following situations were analysed:
- 2 characteristic harbour positions (projecting and indented)
- 2 marina positions (northern Adriatic (Umag) and southern Adriatic (Pelješac)
- 2 widths of harbour entrance (24 and 50 m)
- Without wind and with 4 wind directions (NW, SE, SW and NW)
- 2 winds intensity (2 Bf and 4 Bf)
- 3 culvert positions regarding the marine entrance (in front, left and right)

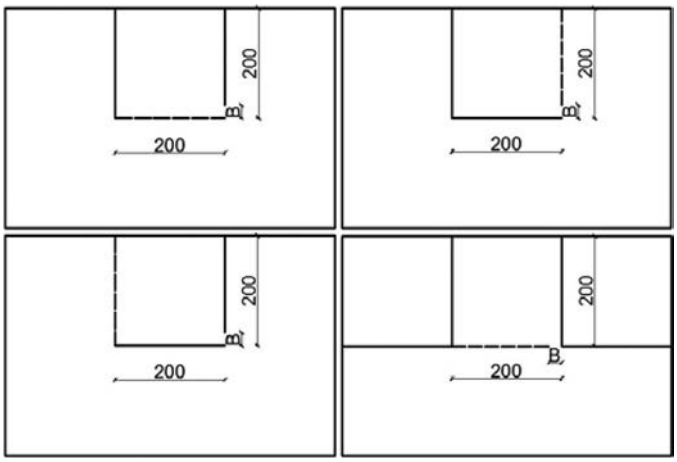


Fig. 1 – Schematic display of projecting marina with culverts in front (top left), culverts to the right (top right), culverts to the left (bottom left) and indented marina with culverts in front (bottom right); B=24/50 m

Sea current generation was implemented with time series of water surface changes along the transect of the open boundary. Hourly sea level dynamics were used on the basis of 7 basic constituent tidal signals (Janeković et al., 2003; Janeković and Kuzmić, 2005; Janeković and Sikirić-Dutour, 2007). The simulation time period was 01/08/2015 to 06/08/2015. Sea temperature and salinity were not taken into account in the model forcing. Unstable forcing by wind with speeds 2 Bf and 4 Bf was used. During the starting (warm-up) period of 24 hours wind speed is 0 and in the following period from 24 to 30 (36) hours rises linearly to the maximum value of 2 Bf (4 Bf). After that the wind remained at a constant value over the course of 36 hours (2 Bf), respectively 24 hours (4 Bf) and then calms linearly over a period of 6 (12) hours. During the rest of the simulation period there is no wind activity (figure 2).

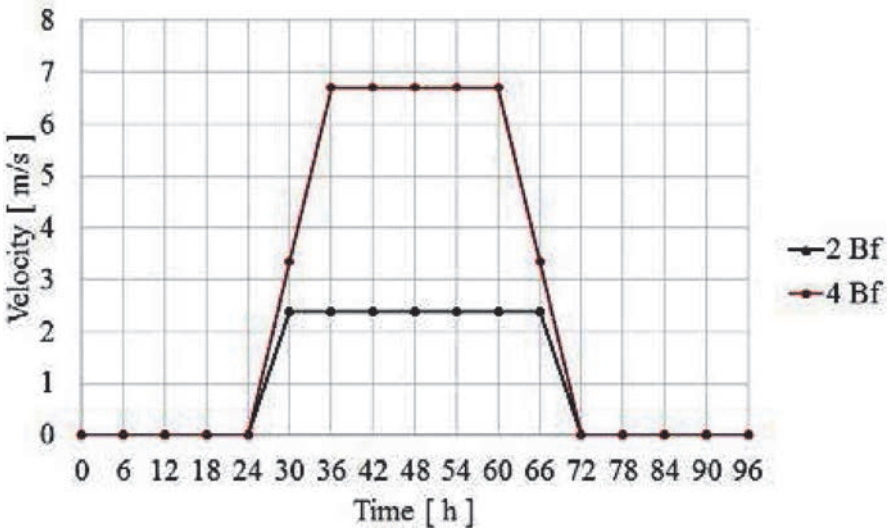


Fig. 2 – Diagram of wind velocity of 2 and 4 Bf over the course of the simulation period

The closure of turbulence model in the model relies on the k – ε formulation (Rodi, 1987) in the vertical direction and Smagorinsky''s concept (1993) in the horizontal direction. Proportional factors for the field of turbulent kinetic energy (TKE) and dissipation (ε) were adopted with values 1 (TKE) and 1.3 (ε) in the horizontal and vertical direction. Bottom roughness and Smagorinsky's coefficient were adopted in the model as spatially homogenous with the values 0.4 and 0.01 m.

2. Results and discussion

Figures 3 and 4 show fields of vertically averaged tracer concentrations for several model simulations. Initial tracer concentration in the enclosed marine aquatorium is a homogenous value of 100. The fields demonstrated are related to the situation with a time shift of 9, 18, 45, 63 and 81 *hour* respectively to the initial condition. Figures 5 and 6 demonstrate the comparison of e-folding times to the situation with and without wind action from NE and SE directions with 2 Bf and 4 Bf intensity for the projecting northern Adriatic marina, the indented northern Adriatic marina, the projecting southern Adriatic marina and the indented southern Adriatic marina.

In figure 5 it is apparent that the optimum seawater exchange situation appears in the projecting marina with culverts placed on the left side. Under wind action there is a consequent significant lessening of seawater retention time in the enclosed part, where the model results point to an equal contribution of winds from SE and NE directions.

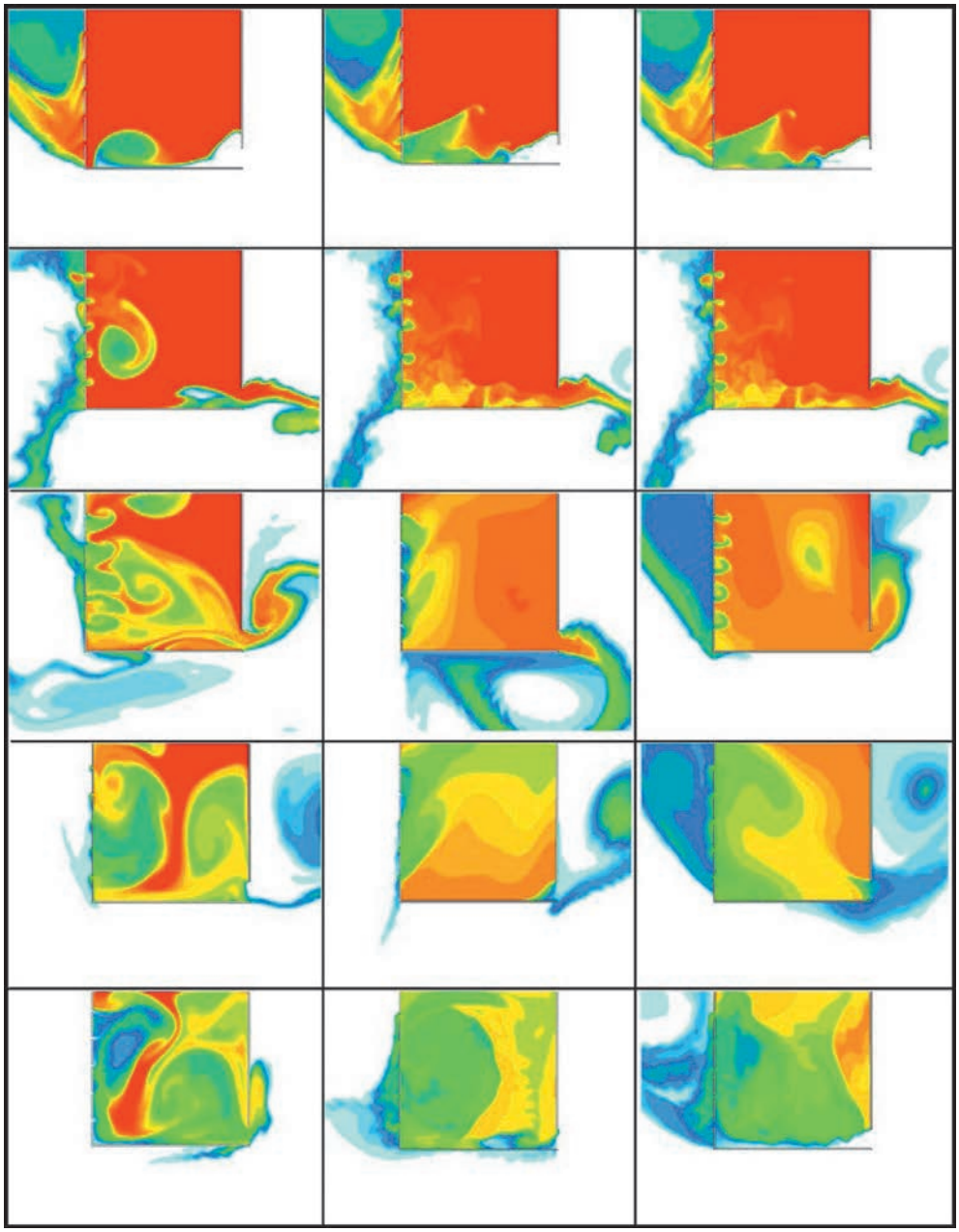


Fig. 3 – Fields of vertically averaged tracer concentrations in the projecting marina in the northern Adriatic with 24 m wide entrance and culverts on the left side (left: variants without wind, centre: variants with NE (direction) wind and 2 Bf intensity, right: variants with SE (direction) wind and 2 Bf intensity; top to bottom: 9th, 18th , 45th , 63th and 81th hour of simulation period

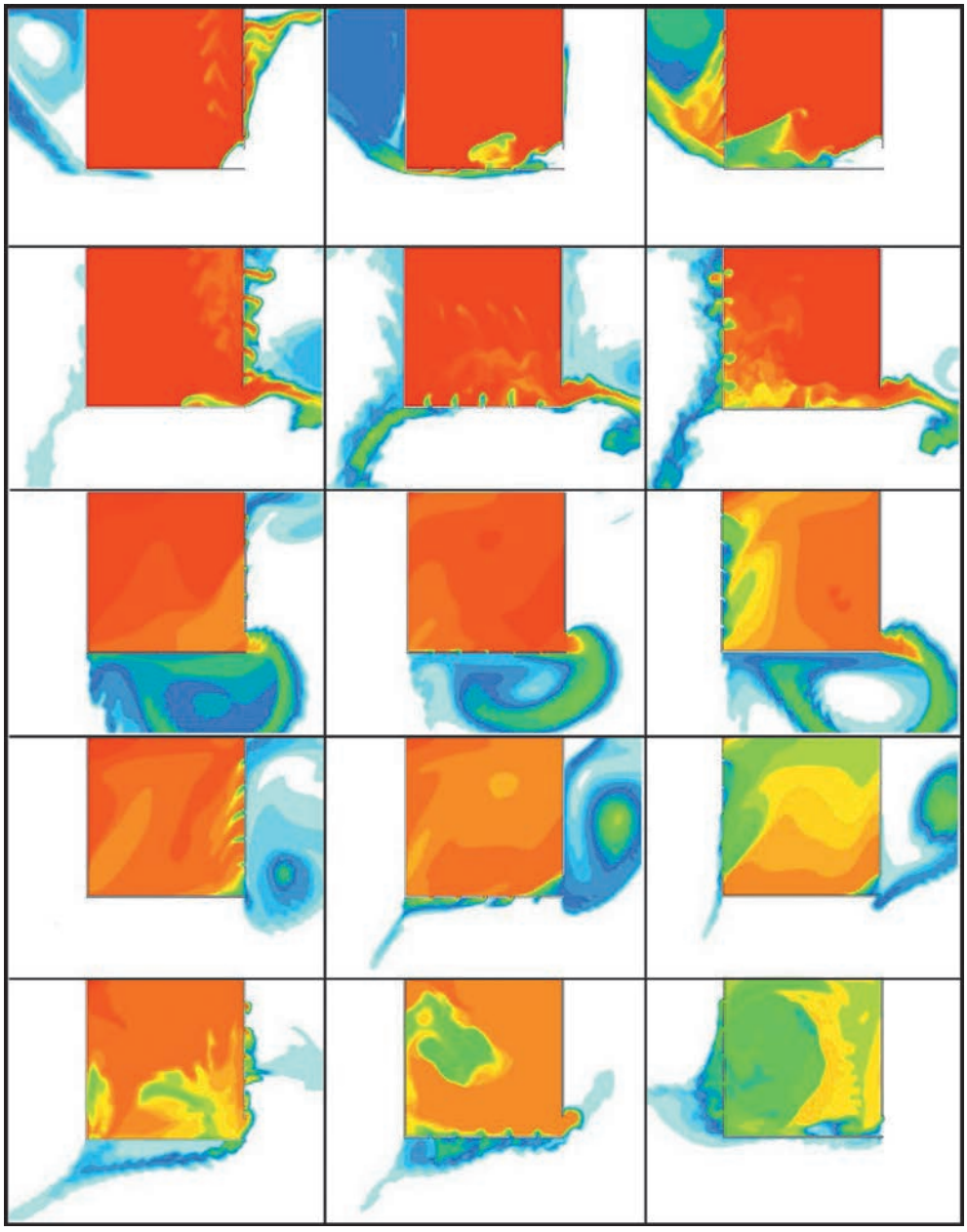


Fig. 4 – Fields of vertically averaged tracer concentrations of the projecting marina in the northern Adriatic with entrance width of 24 m for periods of (blowing) wind from NE direction (and) of 2 Bf intensity (left: variants of marina with culverts on right side, centre: variants of marina with culverts in facing side, right: variants of marina with culverts on left side; top to bottom: 9^{th}, 18^{th} , 45^{th} , 63^{th} and 81^{th} hour of simulation period

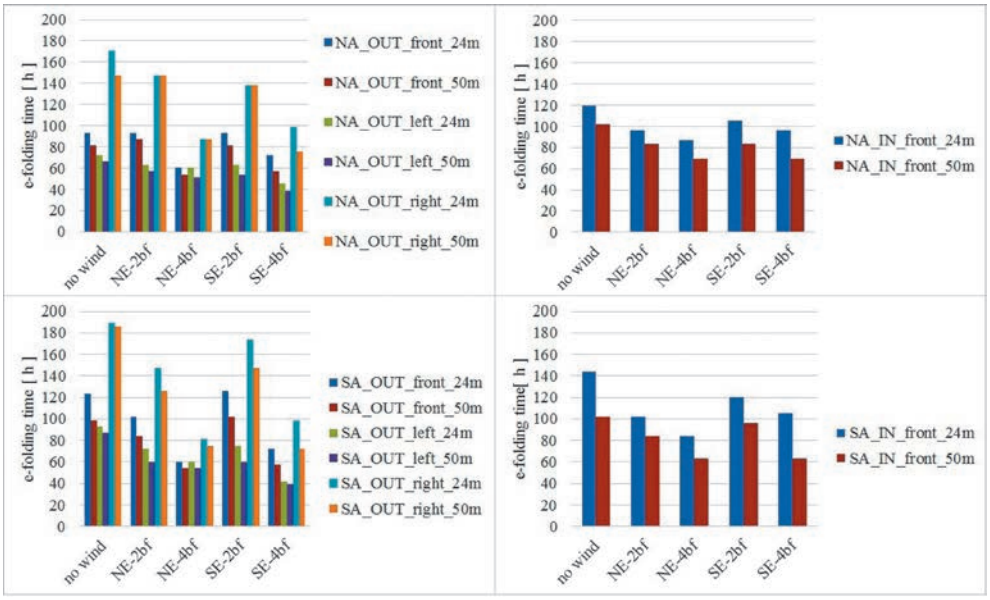


Fig. 5 – Comparison of e-folding times for situations of wind absence and the action of NE and SE (direction) winds with 2 Bf and 4 Bf for the projecting marina in the northern Adriatic (above left), the in-set marina in the northern Adriatic (above right), the projecting marina in the southern Adriatic (below left) and the in-set marina in the southern Adriatic (below right)

For the projecting marina with 24 m wide entrance in conditions where wind was absent, and the culverts placed on front side calculated values of e-folding times amount to 93 h (northern Adriatic), and respectively 123 h (in the case of the southern Adriatic). An increase in entrance width to 50 m results in reduction of e-folding times by 13% (northern Adriatic) and 20% (southern Adriatic). On the other hand, positioning the culvert(s) to the left results in a reduction of e-folding time by 23% (northern and southern Adriatic). Therefore culvert construction on the opposite side of the marina entrance is the best means of ensuring increased seawater exchange.

Wind action contributes significantly to increased seawater exchange. In the projecting marinas equal stimulation for seawater exchange is generated by the influence of SE and NE winds, while in indented marinas NE wind have a greater influence. Under the action of SE and NE winds with 2 Bf intensity e-folding time on average attenuated by 11% (SE) and 17% (NE) in comparison with situations with calm winds, independently of place of culvert construction. In situations of wind action of 4Bf reduction of e-folding time was anticipated more intensively and wind directions SE, NE, NW and SW intensify seawater exchange with pronounced e-folding times on average 40% (SE), 39% (NE), 34% (NW) and 39% (SW in comparison to situations with absence of wind.

It was demonstrated that wind action best contributes to the stimulation of seawater exchange with culverts placed on the right side. In the projecting marina with culverts on the right side under NE and SE direction wind action (2 Bf and 4 Bf) e-folding time on average attenuated by 34% and 32% in comparison with calm wind situations, in the case of the culverts located on the left side attenuation were 24% and 34%, and in the case of the culverts located on the face side attenuation were 24% and 16%.

Considering the influence of tidal signal it was anticipated that in the northern Adriatic, where tidal oscillations are greater, shorter e-folding times would be shown than in the southern Adriatic. In the projecting marinas, when the wind does not blow, e-folding time is 20% shorter in the northern Adriatic in comparison to the south. On the other hand, in the in-set marinas e-folding time is 10% shorter.

4. Tidal-Prism seawater exchange model

As an alternative to the advanced numerical models, for single, relatively small geometry harbour basins, a practical method for determining tracer dispersion with the so-called 'nondimensional' approach is offered with reference to the 'tidal-prism' model on the basis of repeated exchange of tidal volumes (Barber and Wearing, 2002). The model's adopted assumptions are: starting concentration in the aquatorium is uniform, tracer concentration outside the aquatorium is constant and equals 0, during each tidal cycle in the aquatorium a complete admixture occurs, there is no vertical stratification caused by differences in temperature or salinity, sea level oscillation has constant period and amplitude. Water volume in the aquatorium is given in the equation:

$$V = V_m + V_t \cos(\omega t) \tag{1}$$

where: V_m is mean aquatorium volume, V_t is amplitude of tidal volume oscillatory components , t is time and ω is angular velocity defined as $\omega = 2\pi/T$, where T is period of tidal oscillations (daily – T = 24.84 h, half-daily – T = 12.42 h)

Admixture processes differ during the course of ebb and flow cycle. Consequently analysis is divided into different intervals depending on whether the aquatorium is filling or emptying. Governing equations are divided for flood and ebb tide period and are combined over the course of the iterative procedure in order to arrive at an average tracer concentration for the aquatorium after n complete cycles. Except for periodical sea water changes it is possible to include the influence of culverts and of wind through the additional "source" members.

During the ebb tide interval ($0 < \omega t < \pi$), ptracer concentration C and water volume V in the aquatorium meet the following equation:

$$\frac{d}{dt}(CV) = C\frac{dV}{dt} + V\frac{dC}{dt} = QC\,. \tag{2}$$

$$\frac{dV}{dt} = Q + Q_f\,. \tag{3}$$

where: Q is flow through entrance, Q_f is flow through culvert. With solving equation system (2) and (3) for n tidal interval is obtained:

$$C_{e(n)} = C_{f(n-1)} \exp\left\{\frac{-\pi Q_f}{\omega\sqrt{V_m^2 - V_t^2}}\right\} \tag{4}$$

where: $C_{e(n)}$ is tracer concentration at end n of ebb tide interval, and $C_{f(n-1)}$ is corresponding concentration at the end of previous high tide interval

In the case when contribution of wind is considered equation (4) takes the following form:

$$C_{e(n)} = C_{f(n-1)} \exp\left\{\frac{-\pi Q_f}{\omega\sqrt{V_m^2 - V_t^2}}\right\} \exp\left\{\frac{-\pi Q_w \frac{t_w}{T}}{\omega\sqrt{V_m^2 - V_t^2}}\right\} \tag{5}$$

where: Q_w is source member with which contribution of wind action in sea water exchange is simulated, t_w is wind duration.

Over the course of the high tide interval ($\pi < \omega t < 2\pi$), tracer concentration and water volume in the aquatorium meets the following equation:

$$\frac{d}{dt}(CV) = C\frac{dV}{dt} + V\frac{dC}{dt} = bQC\,. \tag{6}$$

$$\frac{dV}{dt} = Q + Q_f\,. \tag{7}$$

where: b is water return parameter leaving during ebb tide (0 – leaving water does not return into the marine aquatorium).

Water volume in the aquatorium at ebb tide amounts to $V_m - V_t$, and volume exchange at high tide is $V_m + V_t$. Resolution of system equation (1), (6) and (7) after n high tide cycle is obtained:

$$C_{f(n)} = C_{f(n-1)} \left[\frac{V_m - V_t}{V_m + V_t} \right]^{(1-b)} \exp \left\{ \frac{-\pi Q_f (1+b)}{\omega \sqrt{V_m^2 - V_t^2}} \right\} \tag{8}$$

In the case when wind contribution is also considered equation (8) takes the form:

$$C_{f(n)} = C_{f(n-1)} \left[\frac{V_m - V_t}{V_m + V_t} \right]^{(1-b)} \exp \left\{ \frac{-\pi Q_f (1+b)}{\omega \sqrt{V_m^2 - V_t^2}} \right\} \exp \left\{ \frac{-\pi Q_w \frac{t_w}{T}}{\omega \sqrt{V_m^2 - V_t^2}} \right\} \tag{9}$$

In the implemetation of calculations the independence of high tide signal influence as well as culvert contribution and wind action is assumed. Consequently in the first step it is necessary to determine the value of parameter *b*. For obtaining the referential values of time concentration, the established time model is added with geometric characteristics described in figure 1 (entrance 24 m and 50 m), but without culverts. On the other hand, equation (9) is applied (only first member with right side), with the condition of achieving sums of the least squared deviations from the numerical values. In the situation of the marinas of the north and south Adriatic, entrances 24 m and 50 m, and for half-daily period of sea level oscillation, the following values were obtained: northern Adriatic and 24 m entrance (b = 0.19), northern Adriatic/ 50 m entrance (b = 0.14), southern Adriatic/24 m entrance (b = 0.60), southern Adriatic/ 50 m entrance (b = 0.48). From the research on the physical model of similar geometric characteristics with marina entrance 50 m (Barber and Wearing, 2002) the value b = 0.14 is obtained at sea level amplitudes from 1 m. The results obtained suggest that the parameter b is primarily a function of high tide oscillation amplitude and not entrance width.

In continuation equations (4) and (8) are used for discovering the values Q_f (culvert contribution), also with the condition of achieving sums of the least squared errors from the concentration values of the obtained corresponding numerical simulation. The last step assumes the application of equation (5) and (9) for discovering the value Q_W (wind contribution).

The procedure was carried out for the northern and southern Adriatic. The tidal amplitude used in the northern Adriatic corresponds to the value of 47 cm and in the southern Adriatic 22 cm. The cases were analysed with daily (24.84 h) and half-daily (12.42 h) oscillation periods. In figures 6 and 7 ‘source’ values Q_f and Q_W are shown obtained by the previously described methodology of the ‘tidal-prism’ model appplication.

In the case of daily tidal oscillation period, culvert influence is the same for the marinas located in the northern and southern Adriatic while in the case of half-dai-

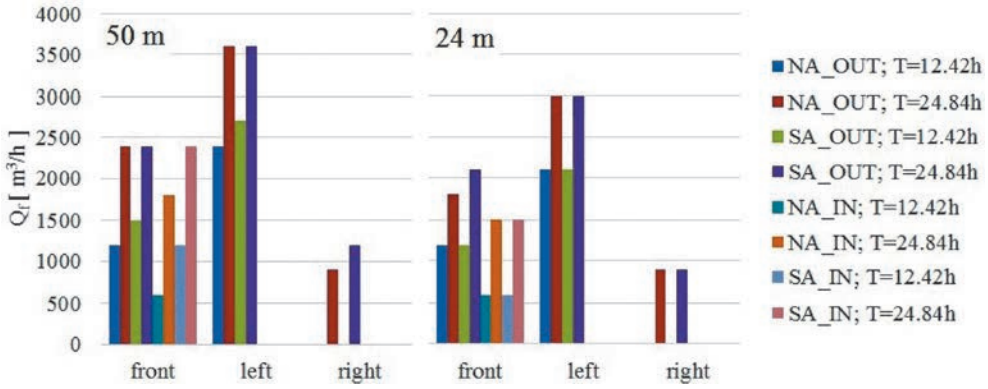


Fig. 6 – 'Source' values Q_f obtained with previously described methodology of the 'tidal-prism' model application

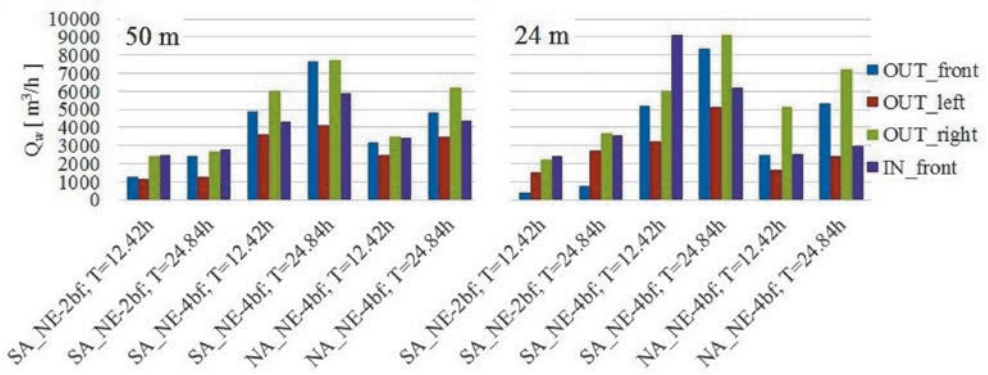


Fig. 7 – 'Source' values Q_w obtained with previously described methodology of the 'tidal-prism' model application (left – entrance width 50 m, right – entrance width 24 m)

ly oscillations a more significant contribution is realized in the southern Adriatic. Such a result is probably the consequence of short-term and small number of tidal cycles in which occur reduced tracer concentration to adopted e-folding values, and because of which a small number of values available for interpolation with the method of the sum of least squares. The influence of wind contribution is more the two and a half times more pronounced than the culvert contribution on average, taking into account all analyzed situations.

5. Conclusions

Research was carried out into seawater exchange in marinas with a 3D numerical and a tidal-prism model. In the conducted analyses the influence of pipe culverts

and wind on seawater exchange between marine aquatoria and the open sea was included. Protective marina breakwaters were considered as vertical constructions. In the analyses carried out marina entrance widths were varied, as well as culvert position, marina topography, location (northern and southern Adriatic) and wind direction and intensity. For a comparative analysis individual calculations were used with so-called 'e-folding' time.

The results demonstrate that seawater exchange is most intensively stimulated by the construction of culverts and breakwaters opposite (left) the entrance and that change in entrance width has less influence on seawater exchange intensity.

Wind action contributes significantly to seawater exchange, out of which the SE and NE directions have approximately equal intensity of influence.

Regarding the influence of tidal oscillation it was anticipated that in the northern Adriatic, where tidal oscillation is greater, e-folding time would be shown to be shorter that in the southern Adriatic. Equally it was shown that seawater exchange is more intensive in projecting marinas.

Within the calculations carried out with the 'tidal-prism' model were included the influence of culverts and wind through activating source members in appropriate process equations. Calculation results indicate that wind has a considerably greater contribution than pipe culverts in seawater exchange.

References

[1] Barber, R.W., Wearing, M.J., (2002): *A mathematical model for predicting the pollution exchange coefficient of small tidal embayments*, Protection and Restoration of the Environment VI, Skiathos, 355-362.

[2] Carević, D., (2015): skripta „Pomorske građevine", Poglavlje 3: Luke nautičkog turizma (Marine)

[3] Cucco, A., Umgiesser, G., (2006): *Modeling the Venice Lagoon residence time*, Ecological modelling, 193, 34-51.

[4] Janeković, I., Bobanović, J., Kuzmić, M. (2003.): The Adriatic Sea M2 and K1 tides by 3D model and data assimilation, *Geophysical Research Abstracts*, 9, 203-217.

[5] Janeković, I., Kuzmić, M. (2005.): Numerical simulation of the Adriatic Sea principal tidal constituents, Ann. Geophys., 23, 3207–3218.

[6] Janeković, I., Sikirić-Dutour. M. (2007.): Improving tidal open boundary conditions for the Adriatic Sea numerical model, *Geophysical Research Abstracts*, 9, 203-217.

[7] Rodi, W. (1987): Examples of calculation methods for flow and mixing in stratified fluids, *Journal of Geophysical Research*, 92(C5), 5305-5328.

[8] Smagorinsky, J. (1993.): *Some historical remarks on the use of nonlinear viscosities*, In: Large eddy simulations of complex engineering and geophysical flows, B. Galperin and S. Orszag (eds.), Cambridge University Press, 1-34.

Scientific paper

Numerical Analysis of Sea Quality in the Northern Adriatic Area

Goran Lončar[1,*], Vladimir Andročec[2,*], Damir Bekić[1,*]

[1]Građevinski fakultet sveučilišta u Zagrebu
[2]Akademija tehničkih znanosti Hrvatske

This study presents the results of a numerical analysis of sea quality in the northern Adriatic area over the course of the period 1997-2001. Boundary conditions on the open border of the model and freshwater confluences were repeated cyclically for fields of temperature and sea salinity in every analysed year. At the interface of the sea and atmosphere in the fields of air temperature, relative humidity, cloud cover and solar radiation cyclically repeated values were also used. The only variable field in the conducted simulations is the field of wind speed which in the period analysed was defined on the basis of results of the numerical atmospheric model Aladin-HR.

Through the conducted research results were obtained which point to the importance of wind dynamics as stochastic parameters over which man does not have a direct influence and in the intensity of the dynamic of chlorophyll-a in the northern Adriatic area. The obtained results demonstrated that the highest concentrations of chlorophyll-a appeared during June and October, and in June 2000 and 2001 increased values were recorded in relation to the remaining years of the analysed period. The reason for this phenomenon is probably the longer retention of nutrient-rich water from freshwater confluences, primarily the river Po, in the surface layer of the sea north of Istria.

The results of the numerical analysis concur with results from an assembly of projects carried out with the aim of defining the causes of frequent forming of mucilaginous aggregates in the northern Adriatic in the course of the period 1997-2001. (Project MAT – Monitoring and study of processes of mucilage formation in the Adriatic and Tyrrhenian Seas, 1999-2002).

Key words:
clorofill-a, Adriatic sea, numerical model

[*]goran.loncar@grad.hr, androcec@hatz.hr, dbekic@grad.hr

1. Introduction

In the northern Adriatic area the formation of mucilaginous aggregates causing algal blooms appears at recurrent periods in intervals of 10-40 years (Fonda Umani et al., 1989), but in recent times their frequency has increased. Accordingly, over the period 1997-2001, the phenomenon appeared in 1997, 2000 and 2001. The phenomenon of mucilage generation comes into being due to synergistic effect of several factors, for example significant changes in concentration of nutrients and their ratios (Degobbis et al., 1999; Fajon et al., 1999), changes of the organic fraction of carbon in the passage through the planktonic nutritional chain (Azam et al., 1999), and prevalences of specific climatological and oceanographic conditions (Degobbis et al., 1995). The northern Adriatic Sea is more productive than, for example, the south-eastern part of the Adriatic sea basin, and in 1997, 2000 and 2001 the formation of mucilage aggregates was recorded during the spring and summer months in the coastal water area of the northern Adriatic. In all three years mucous aggregates formed over the course of May and June and remained until the first part of July, when some episodes of intense wind activity caused the dispersal of the aggregates along the sea column (Precali et al., 2005). Over the course of the other years in the period 1997-2001 this phenomenon was not observed. Vertical and horizontal thermohaline gradients created due to warming of the sea surface layer in May and June as well as intrusions of sea with greater salinity in the intermediate layer, together with retention of water from the river Po, favour the development of mucilaginous blooms in northern Adriatic waters (Russo et al., 2005). Conversely, the western Adriatic current (WAC) transports nutrient-rich sea out of the northern Adriatic area (Grilli et al., 2005) and consequently in some particular years one may expect different periods of retention and/or different concentrations of nutrients to be introduced into the northern Adriatic through freshwater confluences. The western Adriatic current was well-developed in every year marked by an absence of the phenomen of algal bloom. Long-term absence of strong winds also contributes to reduced sea dynamics and the stability of stratification, and the sudden interruption of the algal bloom in 2000 took place immediately after some short episodes of strong winds at the beginning of July (Russo et al., 2005). Despite this, an increase in sea temperature at the surface layer and increased intensity of freshwater confluences in the northern Adriatic area also have a role in the development of mucilage phenomena. For example, greater runoff from the river Po in the period of July coincided with the development of marine algal blooms (Degobbis et al., 2005).

The numerical analyses in this research were conducted with the aim of recognising the particular contribution of the field of wind in sea dynamics, and thereby its influence on the development of the concentration of chlorophyll-a, as an essential parameter of sea quality in the surface layer of the northern Adriatic Sea during the

period 1997-2001. Because of this, in the model simulations every calender year was observed as a separate entity. Therefore, initial and boundary conditions have the following features:

– The calender period 1/1-31/12 was analysed with the same spatial distribution of temperature and sea salinity in initial conditions regardless of the analysed year;
– Spatial and climatic distributions of temperature and sea salinity at the open boundary of the model were defined in the same terms (1/1, 1/4, 1/7 and 1/10) regardless of the analysed year;
– The same climatic series for temperature and air humidity as well as cloud cover for parameterisation of heat exchange with the atmosphere was used, regardless of the analysed year. The same is valid for flow and temperature of freshwater confluences
– The model of sea current is forced on the sea/atmosphere interface with the field of wind which has a variable dynamic over the period of every analysed year in the period 1997-2001.

In chapter 2 the principal elements of parameterisation models and the verification of the results of the numerical model are explained. In chapter 3 the results of the numerical modelling of chlorophyll-a concentration are shown, and in chapter 4 the conclusions of the conducted research are listed.

2. Materials and Methods

Numerical analysis of sea current and dynamic biological-chemical/biochemical parameters were conducted with the three-dimensional numerical model Mike3fm (www.dhigroup.com). The process equations and numerical formulation of the hydrodynamic module in the Mike 3fm model are explained in more detail in Lončar et al. (2011). The model domain covers the area shown in fig, 1. The spatial domain of the numerical model Mike 3fm (final volume method; Sleigh and Gaskel, 1998) is discretised with an unstructured calculation grid which in the horizontal direction has variable space increment Δl from 700m up to 3200m (fig. 2), and in the vertical direction contains 20 sigma layers (Song and Haidvogel, 1994).

Model simulations were initiated on the 1st January in each analysed year over the period 1992-2001. In every year identical fields of temperature and sea salinity at standard oceanographic depths were used (fig. 2) from the oceanographic base data "Dartmouth Adriatic Data Base" – DADB (Galos, 2000). Seasonal fields of temperature and sea salinity were referenced in the following periods: winter (Janu-

ary-March), spring (April-June), summer (July-September) and autumn (October-December) (Cushman-Roisin et al., 2007).

On the open boundary the model uses a free surface time series with 1 hour frequency (Janeković et al., 2003; Janeković and Kuzmić, 2005; Janeković and Sikirić-Dutour, 2007), and fields of temperature and sea salinity in the vertical profile were referenced per Galos (2000) with defined values in the terms 1.January, 1. April, 1. July, 1. March (see fig. 2). Identical spatial distribution of temperature and sea salinity is shown cyclically for every year of the simulation. Freshwater confluences (fig. 1) were parameterised with monthly averaged climatological values of flow and temperature according to Raichich (1996), and salinity was defined with the value 0 PSU. Fields of wind speed, as a boundary condition on the contact of sea and atmosphere, were obtained from the atmospheric model

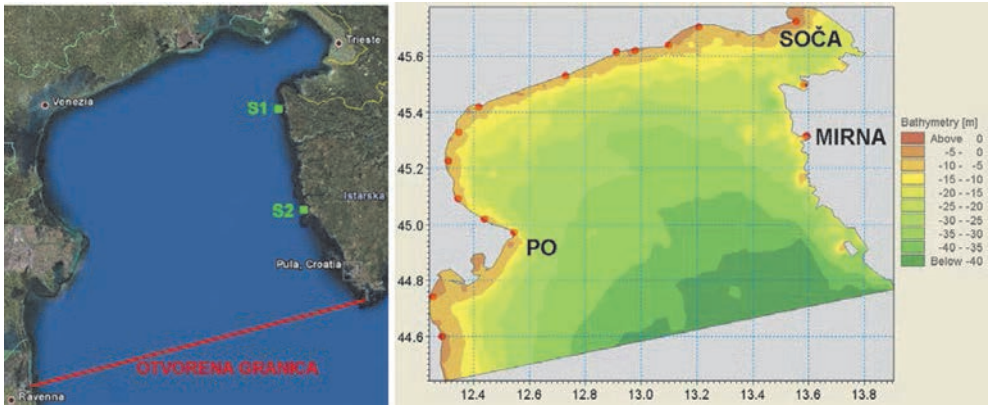


Fig. 1 – Area covered by spatial domain of numerical model Mike 3fm with charted positions ADCP and CTD station (green squares) and position of freshwater confluences in the modeled area

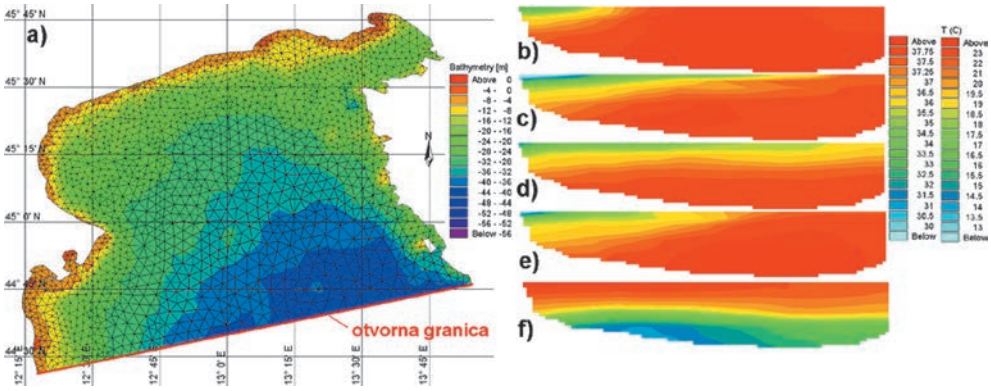


Fig. 2 – Discretisation of the spatial domain of the numerical model Mike 3fm with bathymetric base (a) with fields of salinity (b – winter, c – spring, d – summer, e – autumn) and the field of sea temperature in the vertical profile of the open boundary (f – summer)

Aladin-HR (Members of the ALADIN international team, 1997; Courtier et al., 1991; Cordoneanu and Geleyn, 1998; Brzović and Strelec-Mahović, 1999; Ivatek-Šahdan and Tudor, 2004). For the analysed period data were used with spatial increments of 4 km and temporal steps of 3 hours.

The turbulence model relies on the k-ε formulation (Rodi, 1987) in the vertical direction and the Smagorinsky concept (1993) in the horizontal direction. Adopted values of coefficients in the hydrodynamic module of the Mike 3fm model are the following: proportionality factors for scalar fields turbulent dispersion of temperature and salinity $Pr_{V\text{-}D} = 0.05$ (vertical direction) and $Pr_{H\text{-}D} = 0.15$ (horizontal direction); proportionality factors for turbulent kinetic energy (TKE) and dissipation *(ε)* $Pr_{V\text{-}k\varepsilon} = 1.3$ (vertical direction) and $Pr_{H\text{-}k\varepsilon} = 1.0$; roughness = 0.01m; Smagorinsky coefficient = 0.2; coefficient of wind friction 0.0016 (Wu, 1994); coefficient in Angstrom's law $a = 0.25$ and $b = 0.52$; coefficient in Dalton's law of evaporation = 0.9; coefficient of absorption of light energy in the surface layer in Beer's law = 0.2; coefficient of light attenuation in Beer's law = 0.95.

Verifications of the results of the numerical model were carried out on the basis of measurements of results obtained in the project: "Adriatic sea monitoring programme" (Andročec et al., 2009). Measurements of current, temperature and sea salinity were carried out over the period 11/2007-11/2008. As the primary interest of this research is the implementation of analysis in the period 1997-2001, for which measures of current were not available, a previously described numerical model was applied for the period 1/1/2008-15/11/2008 as well, along with the adoption of the same incipient and boundary conditions as in simulations for the period 1997-2001. Data on the variability of the field of wind for the period of simulation 1/1/2008-15/11/2008 were taken from the results of the numeric model Aladin-HR. Measurements of sea current were achieved with ADCP current measurements at the S1 and S2 stations (fig. 1). A comparison of the measured and modeled hour averaged current speed at 8m depth for the period 15/6/2008-15/7/2008 is shown in figure 3. Error of the modeled values of hourly averaged current components at the positions of the current measuring stations S1 and S2 in relation to measured values for the period of simulation 15/6/2008-15/7/2008 acquired the values (for the root mean square error) of RMSE = 0.057 (for *u* component) and RMSE=0.074 (for *v* component). In figure 4 is shown the comparison of the modeled and measured distributions of temperature and salinity of the sea in the vertical flow measuring stations S1 and S2 in the terms 13/3/2008 and 27/6/2008.

The hydrodynamic module of the Mike 3fm model gives a three-dimensional picture of the current required for a further course of calculations of concentrated parameters of sea quality. For this the module ECOlab was used, which acquires information about flow dynamics directly from the hydrodynamic module

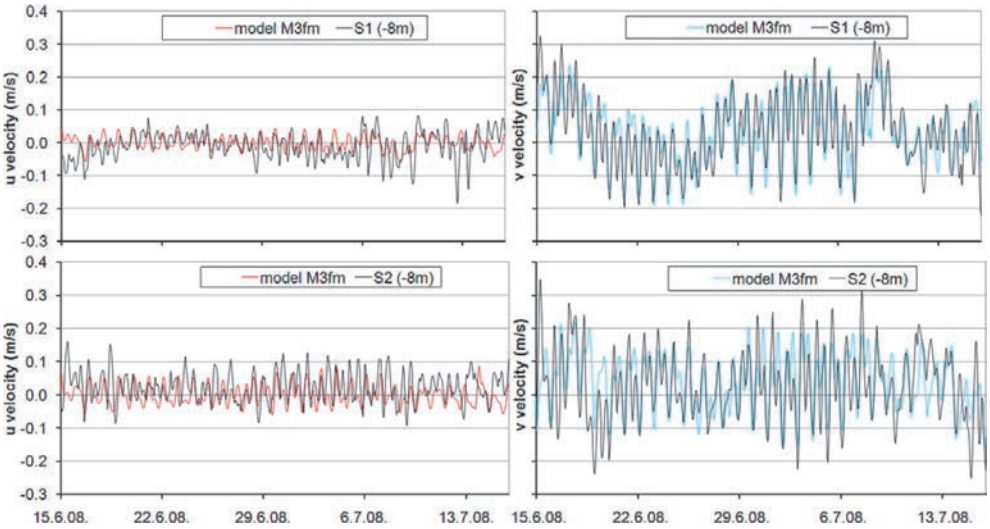


Fig. 3 – Comparison of measured and modeled (Mike 3fm) hourly averaged current speed at position of current measuring station S1 (above) and S2 (below) at 8m depth

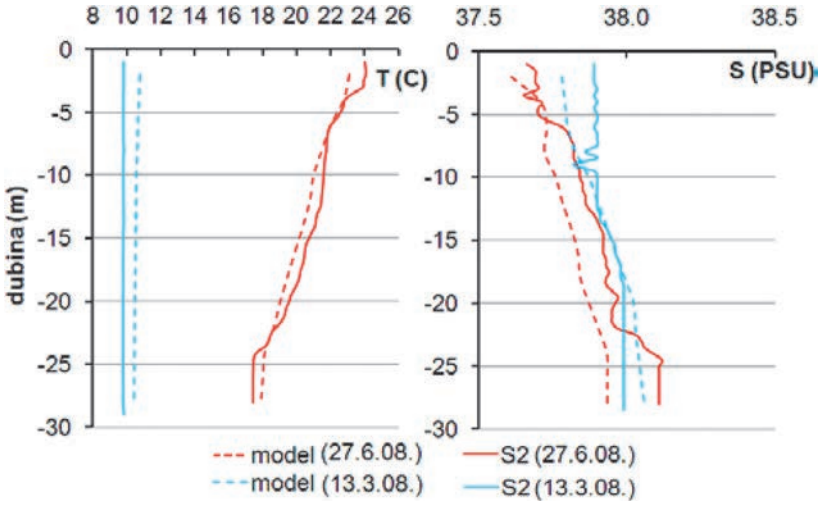


Fig. 4 – Comparison of modeled and measured distributions of temperature and sea salinity along water column at current measuring stations S1 and S2 in terms 13/3/2008 and 27/6/2008

and carries out further treatment of convectively-dispersed transport with bio-chemical reactions. In the analysis of sea quality numerical formulation was used which treats the concentration of variables: dissolved oxygen *(DO)*, ammonia *(NH_4)*, nitrite *(NO_2)*, nitrate *(NO_3)*, orthophosphate *(PO_4)* and chlorophyll-a *(CHL)* as process variables. For each of the process variables is formed an appropriate differential equation, describing the rate of its changes. It is based on the processes which take place within observed ecosystems. In the applied approach, except for

the process variables, constants, forcing functions and auxiliary variables were used.

For analysis of the transportation of scalar fields the 3D quickest-sharp scheme was used which is especially suitable for cases of pronounced gradients and is affiliated to the so-called CWC scheme (Consistency with Continuity). For integration, the Runge-Kutta method (4^{th} order) was used. Reaeration is a process which describes the interchange of oxygen between dissolved oxygen in water and the atmosphere, and the applied term includes a saturation level of oxygen in water that depends on salinity and temperature. A process of rapid reaeration is correlated with rapid wind, rapid current and deep water. The process of photosynthetic production of oxygen is described in a relative relationship against given maximum production and is variable in time with the relative duration of any particular day. For the value of maximum production of oxygen through the process of photosynthesis the base value $1.5gO_2/m^2/day$ was adopted. Photosynthetic production of oxygen and respiration of autotrophs is variable by depth because of the correlation of light intensity with autotroph concentration. Variation in depth function is described by Lambert-Beer's law in which to the coefficient of light attenuation is adopted the value 15m (GF, 2010). The respiration of autotrophs and heterotrophs is dependent on oxygen and is correlated with sea temperature. To a rate of respiration at 20^0C is applied a constant value in the course of one calendar day from $0.2gO_2/m^2/day$. The temperature coefficient for respiration is applied with value 1.08, and the concentration of dissolved oxygen for which the process rate reduces to 50% as a result of lack of oxygen is defined with the value 2mg/l. Sediment oxygen demand is hypothesised as being correlated with the concentration of oxygen and temperature via Michaelis-Menten kinetics so an expression is used with the rate $0.2gO_2/m^2/day$, Arhenius' temperature coefficient 1.07 and the concentration of oxygen at 50% saturation 2mg/l. Change of oxygen due to nitrification is defined as a process of the first order given ammonium and the Michaelis-Menten process and given oxygen with adopted values $3.42gO_2/gNH_4$-N (requirement for oxygen for nitrification $NO_2 \rightarrow NO_3$) with appropriate adopted value of Arhenius' temperature coefficient 1.088 and concentration of oxygen at 50% saturation from 2mg/l. The ammonium balance includes ammonium reception obtained from the transformation of ammonium into nitrite and uptake by algae and bacteria. The process of transformation of ammonium into nitrite is hypothesised as a process of the first order regarding the concentration of ammonia in direct conversion, and is reliant on temperature and concentration of oxygen. Consumption of ammonium by algae is referenced to photosynthesis and respiration and as a constant value per gram of net production of oxygen. During the night, consumption of ammonium remains constant. During the course of the day, intensity of consumption is reliant on the concentration of ammonium and that is defined by Michaelis-Menten kinetics with an adopted constant value 0.066gN/gDO for estimated quantity NH_4-N consumed by algae. A function of limitation of nutrients was introduced in which is half saturation con-

centration for nitrogen adopted with 0.05, and for phosphorus with the value 0.01. Uptake of ammonium on the part of bacteria is also described by Michaelis-Menten kinetics in which constant consumption is adopted with 0.109gN/gDO, and half saturation concentration with 0.05mg/l. Denitrification, through which nitrates transform into free atmospheric nitrogen, is interpreted as a process of the first order reliant on temperature. To a rate of nitrification to 20^0C the value 0.1 1/day was adopted, with adopted Arhenius' temperature coefficient 1.16. Phosphorus is consumed in the process of photosynthesis. Taking into account the uptake of phosphorus in the production of algae the balance equation used for concentration of orthophosphates includes consumption of phosphorus by algae (referenced to photosynthesis and respiration) and that as a constant value per net gram production of oxygen. In the course of a day the intensity is reliant on the concentration PO_4 which is defined by Michaelis-Menten kinetics, in which quantity PO_4 consumed by algae is adopted with the constant 0.0091gP/gDO. Uptake of phosphorus by bacteria upon mineralisation of organic substances is also described by Michaelis-Menten kinetics with the adopted value 0.015gP/gDO and 0.05mg/l for half saturation concentration. The production of chlorophyll-a is hypothesised as proportional to phytoplanktonic production of carbon and thereby the production of oxygen. The balance equation for chlorophyll-a concentration is based on the ratio chlorophyll-a/carbon (adopted 0.025mgCHL/mgC) and carbon/oxygen for primary production (adopted 0.2857mgC/mgO). The die-off and sedimentation rate of chlorophyll-a were defined with values 0.01 1/day and 0.2m/day respectively.

The initial values of the model parameters were given as homogenous, based on data measured at the oceanographic station situated at 1nM from the Ruđer Bošković Centre for Sea Research in Rovinj (continued tracking in the period 1975-2005). The values are representative of the Rovinj aquatorium and certainly deviate from reference values for the western coastline of the northern Adriatic, but due to the primary aim of this research the used data are relevant. Therefore, adopted values for initial conditions amounts to: CHL = 0.00075mg/l, PO_4 = 0.0024mg/l, NO_2 = 0.0002mg/l, NO_3 = 0.08mg/l, NH_4 = 0.00116mg/l i DO = 7.2mg/l. Further, because of the data shortfall on all parameter concentrations on the open boundary (effect of open sea) in the course of the simulation period it was necessary to use the values obtained in the measurements at the oceanographic station at 1nM from Rovinj. Concentrations of parameters in locations of freshwater confluences were adopted with constant values in the course of the simulation period and presented in table 1. For the rivers Mirna, Dragonja and Soča the same concentrations were used, obtained averaging values from 2006, 2009 and 2010 (Hrvatske vode, 2010; 2011). For every other river of the Italian watershed (picture 1) concentrations were used which match annual averages for the river Po (Wit and Bendoricchio, 2001; Palmeri et al, 2005).

Table 1 – Concentrations of parameters DO, NH_4, NO_2, NO_3, PO_4 i CHL=0 mg/l in places of freshwater discharges (rivers)

	Mirna			Stella-Reno	Mirna, Dragonja, Soča
	2006	2009	2010	adopted	adopted
DO (mg/l)	8.8	8.2	9.3	9.5	8.8
NH_4 (mg/l)	0.083	0.024	0.029	0.074	0.045
NO_2 (mg/l)	0.013	0.009	0.003	0.011	0.008
NO_3 (mg/l)	1.509	1.264	0.726	2.240	1.166
PO_4 (mg/l)	0.011	0.008	0.008	0.075	0.009

It is worth noting that the numerical analyses were conducted with the purpose of quantifying the development intensity of chlorophyll-a and serve only for comparison of individual situations over the period 1992-2001. Accordingly, the sensitivity analysis of the constants used and comparison with measured values of modeled parameters of sea quality were not conducted.

3. Results and Discussion

In figure 5 are shown monthly averaged fields of current at a depth of 1m for June and July 1997, 1999, 2000 and 2001.

In figure 5 one may observe the absence of outgoing western Adriatic current in the surface layer during the months of June 2000 (fig. 5e) and 2001 (fig. 5g). The most pronounced inflow current is noted in 2000, and it is somewhat weaker in 2001 and 1997. It is an interesting fact that the highest concentrations of mucilage correspond exactly with these events in 2000 and with somewhat weaker intensity in 2001 and 1997. Furthermore, in 1999, during the course of June (fig. 5c) a well developed western Adriatic current was present, responsible for the removal of surface water rich in nutrients from the northern Adriatic. In that year no mucilage phenomenon was recorded. Over the course of July in each year a development of intensive western Adriatic current took place, which also corresponds with term of disappearance of mucilage in 1997, 2000 and 2001.

In figure 6 are shown monthly averaged fields of concentration of chlorophyll-a at a depth of 1m for the months of June and July in the years 1997, 1999, 2000 and 2001.

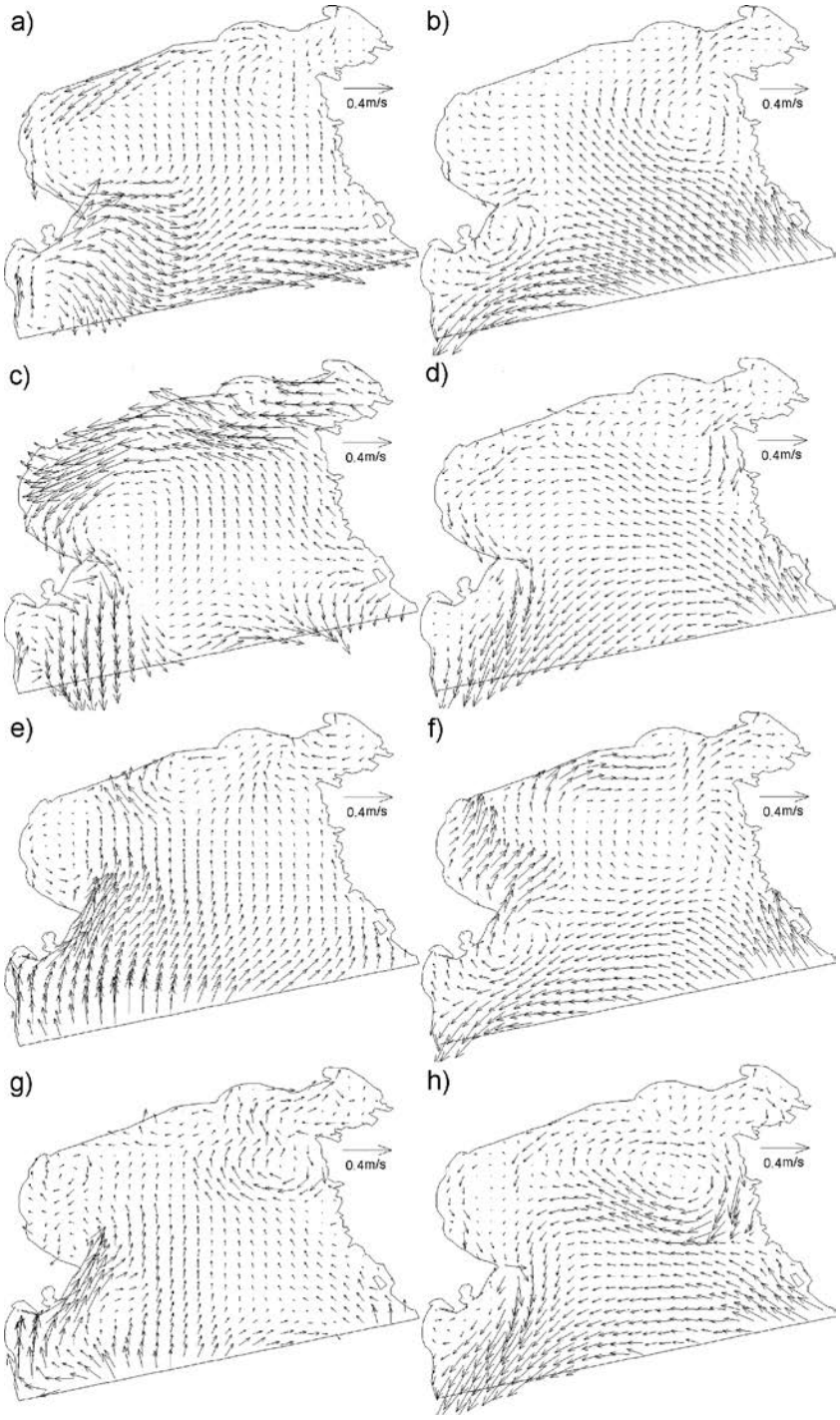


Fig. 5 – Monthly averaged fields of current at a depth of 1m for June (left) and July (right) 1997 (a,b), 1999 (c,d), 2000 (e,f), and 2001 (g,h)

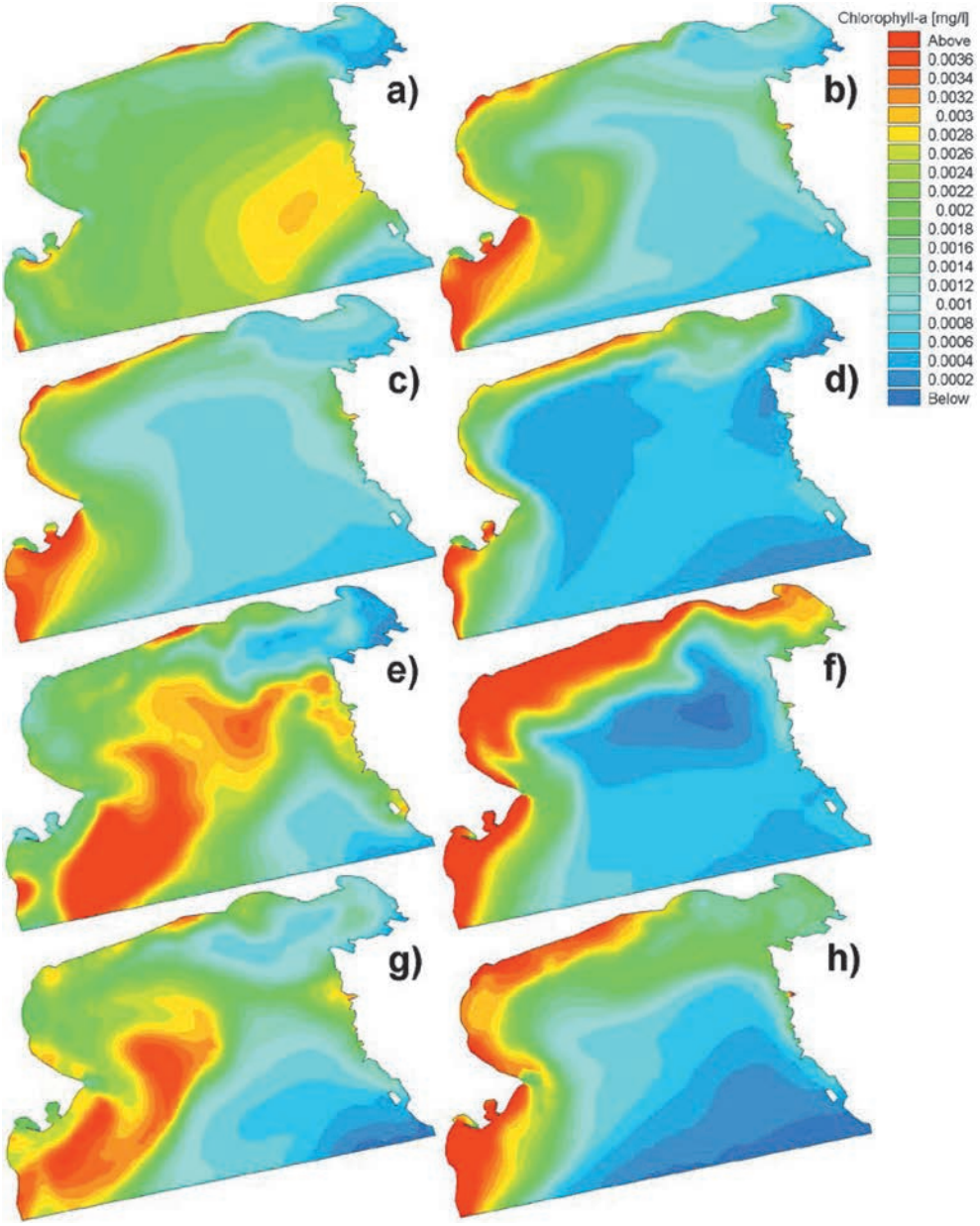


Fig. 6 – Monthly averaged fields of concentration of chlorophyll-a at a depth of 1m for June (left) and July (right), 1997 (a,b), 1999 (c,d), 2000 (e,f) and 2001 (g,h)

With figure 6 may be observed the appearance of maximum concentrations of chlorophyll-a during the month of June 2000, when the greatest concentration appeared on the Istrian coast. After that, over the course of July, concentrations of algal blooms arrived, of greatest concentration in the west and north part of the analysed

area. A similar situation developed in 1997 and 2000 with a difference in the position of maximum concentration and spatial distribution of chlorophyll-a in the surface layer. Namely, in June 1997 the field of concentration is more homogenous than in 2001, and the greatest concentration appeared in the eastern part of the observed area. In 1999, a more appreciable increase of concentration of chlorophyll-a in the central or eastern part of the modeled area (northern Adriatic) was not registered. Independently of the analysed years, in the course of July, outflow of the western Adriatic current became stronger, leading to significant reduction of chlorophyll-a in the northern Adriatic. These results are in agreement with results of conducted monitoring within the MAT project.

In figure 7 are shown time series of modeled concentrations of chlorophyll-a in the surface layer of the sea with monthly averaged values for 1999, 2000 and 2001 at the position of measuring station S2 (fig.1). For comparison are shown values of average seasonal model of chlorophyll-a concentration obtained on the basis of continued measurements at the same position in the period 1972-2005. (Andročec et al.; www.mzopug.hr/doc/Zastita_okolisa)

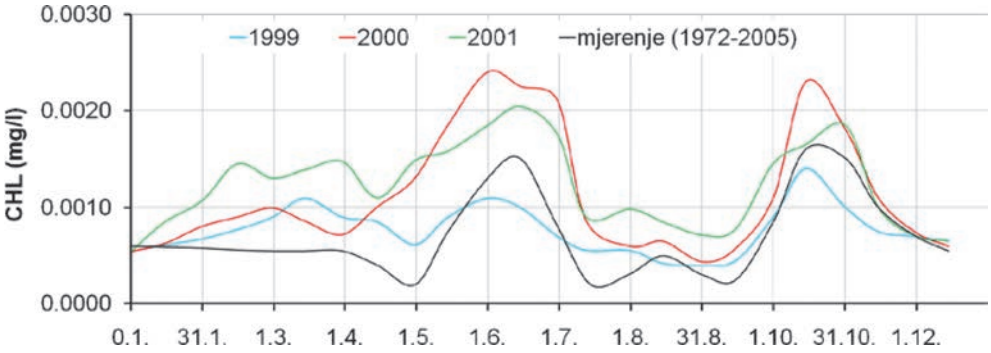


Fig. 7 – Time series of modeled concentrations of chlorophyll-a in the surface layer of the sea with monthly averaged values for 1999, 2000 and 2001 at the position of measuring station S2 and values of average seasonal models of concentration of chlorophyll-a obtained on the basis of continued measuring at the same position in the period 1972-2005

With figure 7 it seems that the model values follow the trend of the average seasonal model based on measured values. In 1999 the maximum concentration was calculated in the middle of October, and not in May. In 2000 and 2001 over the course of April and May came a sudden increase in the concentration of chlorophyll-a. Increased concentration maintains itself through the month of June and in the first part of July comes to a sudden drop in concentration due to development of the western Adriatic current and a general amplification of sea dynamic. It is worth noting that the modeled terms of a sudden decrease in chlorophyll-a concentration in July 2000 and 2001 coincides with the registered terms of interruption of algal blooms in the MAT project (Precali et al., 2005).

4. Conclusions

Research was conducted on the impact of the dynamic field of wind on the dynamic concentration of chlorophyll-a in the northern Adriatic area. Research was conducted with a three-dimensional model of current, using a hydrodynamic and ecological module. Current obtained by calculation with the hydrodynamic module base is a further estimate of the concentration of sea quality parameters, including the concentration of chlorophyll-a. Model results of current as well as temperature and sea salinity were verified by measurements in situ at current measurement and CTD stations.

In the implementation of modelling identical and cyclical repeated initial and boundary conditions for every single year from period of analysis 1997-2001 were used. The only difference in the forcing of the model is spatial and time variability in the field of wind speed, in which the results of numerical atmospheric model Aladin-HR for the period 1997-2001 were used.

The model results of current for 1998 and 1999 point to a well developed western Adriatic outflow, while in the other years considerably weakened outflow of the sea from the northern Adriatic area via the western Adriatic current is noted. Retention of water from freshwater confluences in the modeled area of the northern Adriatic is especially present in the course of June 1997, 2000 and 2001. In June 2000 the maximum concentration of chlorophyll-a appeared in comparison with the other months and years from the analysed time span. The term of appearance of maximum concentration of chlorophyll-a, obtained with the numerical model, coincides with the term of intensive formation of mucilage aggregates in 2000. With this, it is possible to confirm the hypothesis in which the field of wind is considered an important factor in the creation of suitable conditions for the formation of mucilage aggregates in the northern Adriatic area.

Acknowledgements

The authors would like to acknowledge the assistance of the National Hydrometeorological Servise, for the research agreement and collaboration with Faculty of Civil Engineering regarding data obtained by the prognostic model ALADIN for the period 1999 to 2001.

References

[1] Andročec, V., Beg-Paklar, G., Dadić, V., Djakovac, T., Grbec, B., Janeković, I., Krstulović, N., Kušpilić, G., Leder, N., Lončar, G., Marasović, I., Precali, R., Šolić, M.: The Adriatic Sea Monitoring Program – Final Report, Zagreb, 2009.

[2] Azam, F., Fonda Umani, S., Funari, E. (1999.): Significance of bacteria in the mucilage phenomenon in the northern Adriatic Sea, *Ann. Ist. Sup. Sanita.*, 35, 411-419.

[3] Brzović, N. (1999.): Factors affecting the Adriatic cyclone and associated windstorms, *Contributions to Atmospheric Physics*, 72, 51-65.

[4] Brzović, N., Strelec-Mahović, N. (1999.): Cyclonic activity and severe jugo in the Adriatic, *Physics and Chemistry of the Earth (B)*, 24, 653-657.

[5] Courtier, P.C., Freydier, J.F., Geleyn, F., Rochas, M. (1991.): The ARPEGE project at METEO-FRANCE, *Proceedings from the ECMWF workshop on numerical methods in atmospheric models*, 193-231.

[6] Cordoneanu, E., Geleyn, J.F. (1998.): Application to local circulation above the Carpathian-Black Sea area of a NWP-type meso-scale model, *Contributions to Atmospheric Physics*, 71, 191-212.

[7] Cushman-Roisin, B., Korotenko, K., Galos, C., Dietrich, D. (2007.): Simulation and characterization of the Adriatic Sea mesoscale variability, *J. Geophys. Res.*, 112, C03S14, doi:10.1029/2006JC003515.

[8] Degobbis, D., Fonda Umani, S., Franco, P., Malej, A., Precali, R., Smodlaka, N. (1995.): Changes in the northern Adriatic ecosystem and appearance of hypertrophic gelatinous aggregates, *Sci. Total Environ.*, 165, 43-58.

[9] Degobbis. D, Malej, A., Fonda Umani, S. (1999.): The mucilage phenomenon in the northern Adriatic Sea. A critical review of the present scientific hypotheses, *Ann. Ist. Sup. Sanita.*, 35, 373-83.

[10] Degobbis., D., Precali, R., Ferrari, C.R., Djakovac, T., Rinaldi, A., Ivančić, I. (2005.): Changes in nutrient concentrations and ratios during mucilage events in the period 1999.-2002. *Sci. Total Environ.*, 353, 103-14.

[11] Fajon, C., Cauwet, G., Lebaron, P., Terzić, S., Ahel, M., Malej, A. (1999.): The accumulation and release of polysaccharides by planktonic cells and the subsequent bacterial response during a controlled experiment, *FEMS Microb. Ecol.*, 29, 351-63.

[12] Fonda Umani, S., Ghirardelli, E., Specchi, M. (1989.): *Gli episodi di „mare sporco" nell'Adriatico dal 1729 ai giorni nostri*, Regione Autonoma Friuli Venezia Giulia, Direzione Regionale dell'Ambiente, Trieste, 178 str.

[13] Galos, C. (2000.): *Seasonal circulation in the Adriatic Sea*, M.S. thesis, Dartmouth Coll., Hanover, N.H., 127str.

[14] Grilli, F., Paschini, E., Russo, A., Precali, R., Supić, N. (2005.): Circulation and horizontal fluxes in the Northern Adriatic Sea in the period June 1999. – July 2002. Part I: geostrophic circulation and current measurement, *Sci. Total Environ.*, 353, 57-67.

[15] Hrvatske vode (2010.): *Kakvoća površinskih voda u RH u 2009. godini*, Zagreb.

[16] Hrvatske vode (2011.): *Kakvoća površinskih voda u RH u 2010. godini*, Zagreb.

[17] Ivatek-Šahdan, S., Tudor, M. (2004.): Use of high-resolution dynamical adaptation in operational suite and research impact studies, *Meteorol. Z.*, 13, 99-108.

[18] Janeković, I., Bobanović, J., Kuzmić, M. (2003.): The Adriatic Sea M2 and K1 tides by 3D model and data assimilation, *Geophysical Research Abstracts*, 9, 203-217.

[19] Janeković, I., Kuzmić, M. (2005.): Numerical simulation of the Adriatic Sea principal tidal constituents, Ann. Geophys., 23, 3207–3218.

[20] Janeković, I., Sikirić-Dutour. M. (2007.): Improving tidal open boundary conditions for the Adriatic Sea numerical model, *Geophysical Research Abstracts*, 9, 203-217.

[21] Lončar, G., Beg-Paklar, G., Janeković, I. (2011.): Influence of density stratification on effluent plume dynamics, Oceanologia, 53(2), 565-585.

[22] Palmeri, L., Bendoricchio, G., Artioli, Y. (2005.): Modelling nutrient emissions from river systems and loads to the coastal zone: Po River case study – Italy, *Ecological Modelling*, 37–53.

[23] Precali. R., Giani, M., Marini, M., Grilli, F., Pečar, O., Paschini, E. (2005.): Mucilage events in the northern Adriatic in the period 1999-2002: typology and distribution, *Sci. Total Environ.*, 353, 10-23.

[24] Rodi, W. (1987.): Examples of calculation methods for flow and mixing in stratified fluids, *Journal of Geophysical Research*, 92(C5), 5305-5328.

[25] Russo, A., Maccaferri, S., Djakovac, T., Precali, R., Degobbis, D., Deserti, M. (2005.): Meteorological and oceanographic conditions in the northern Adriatic Sea during the period June 1999.-July 2002.: influence on the mucilage phenomenon, *Sci. Total Environ.*, 353,24-38.

[26] Sleigh, D. H., Gaskel, P.H. (1998.): An unstructured finite volume algorithm for predicting flow in rivers and estuaries, *Computer and Fluids*, 27(4), 479-508.

[27] Smagorinsky, J. (1993.): *Some historical remarks on the use of nonlinear viscosities*, In: Large eddy simulations of complex engineering and geophysical flows, B. Galperin and S. Orszag (eds.), Cambridge University Press, 1-34.

[28] Song, Y., Haidvogel, D. (1994.): A semi-implicit ocean circulation model using a generalised topography-following coordinate system, *Journal of Comp. Physics.*, 115, 228-244.

[29] Wit, M., Bendoricchio, G. (2001.): Nutrient fluxes in the Po basin *Sci. Total Environ.*, 273, 147-161.

[30] Wu, J. (1994.): The sea surface is aerodynamically rough even under light winds, *Boundary layer Meteorology*, 69, 149-158.

[31] www.mzopug.hr/doc/Zastita_okolisa/Zastita_od_oneciscenja_voda_izvjesce.pdf

Scientific paper

Analysis on Durability of Reinforced Concrete Bridges in Maritime Environment

Marija Kušter Marić[1,*], Jure Radić[1,*], Joško Ožbolt[2,*], Gojko Balabanić[3,*], Filip Oršanić[4,*], Zlatko Šavor[1,*]

[1]University of Zagreb – Faculty of Civil Engineering
[2]University of Stuttgart – Institute of Construction Materials
[3]University of Rijeka – Faculty of Civil Engineering,
[4]Sofistik – Munich, Germany

The paper provides a review of durability issues of the Adriatic large concrete arch bridges built in two periods: in the 1960's and 1970's and at the last turn of the century. Inadequate attention to durability issues in combination with very aggressive maritime environment has led to rapid deterioration of the structures caused by corrosion of reinforcement steel in concrete.

In order to develop consistent maintenance policy that would provide efficient and effective management of bridges in maritime environment, it is necessary to determine remaining service life of the structures. In this purpose the coupled 3D chemo- hygro-thermo mechanical model is developed and implemented into FE code. The applications of the model are illustrated on three numerical examples. The numerical results are in good agreement with the available experimental data leading to the conclusion that the model is able to realistically predict service life of reinforced concrete structures exposed to chlorides and mechanical damage results.

In service performance of the Adriatic large concrete arch bridges shown and numerical analysis confirmed the following: (i) cracks in concrete cover, which are the consequences of mistakes in conceptual design as well as errors and negligence on site, significantly reduces depassivation time of reinforcement bar; (ii) corrosion rate is much higher in poor than in good quality concrete; (iii) the maximal numerical results of corrosion rate, achieved at critical

[*]marijak@grad.hr, jradic@grad.hr, ozbolt@iwb.uni-stuttgart.de, mailto:marijak@grad.hr gojko@gradri.hr, filip.orsanic@iwb.uni-stuttgart.de , savor@grad.hr

water saturation, are within the limits for concrete in splash zone, as one of the most critical environment for corrosion of steel in concrete, (iv) corrosion induced cracks significantly reduce pull-out capacity of reinforcement.

Key words:

Concrete arch bridges, corrosion, chloride, cracks, modelling

1. Introduction

Bridges are a vital part of any transportation network. Typically they comprise just a few percent of the total length, but their share in the overall value of the network is approximately ten times higher. Ever increasing degradation of structures and inevitable increase in requirements posed on bridges during their service life is main challenge in maintenance of bridges worldwide.

The older Adriatic large concrete arch bridges suffered greatly over decades of service, due to combination of aggressive maritime conditions (high winds, chloride ingress, freeze/thaw and thermal action) and inadequate attention to durability issues. Serious deterioration of structural members with reinforcement corrosion being the major issue, led to many complex and expensive repairs. In order to develop consistent maintenance policy that would provide efficient and effective management of Adriatic concrete arch bridges, it is necessary to determine remaining service life of the structures.

Prediction of service life of new or already damaged reinforced concrete structures requires a reliable numerical model, which can realistically simulate major degradation processes in reinforced concrete structures. The major cause of deterioration of reinforced concrete structures in maritime environment is chloride-induced corrosion, therefore it is important to have a numerical tool, which is able to realistically simulate corrosion processes before and after depassivation of steel in concrete and the consequences for the structure. Moreover, by employing such a model, it is possible to formulate simple engineering models and design rules in order to increase the durability of structure and reduce its maintenance costs. In this purpose the coupled 3D chemo-hygro-thermo mechanical model is developed and implemented into FE code in order to enable a realistic prediction of service life of reinforced concrete structures exposed to chlorides and mechanical damage. The applications of the model are illustrated on three numerical examples and compared with available experimental results.

2. In-service performance of the Adriatic concrete arch bridges

Seven large reinforced concrete arch bridges were constructed over the past five decades with spans ranging from 140 m to almost 400 m (Table 1). The Adriatic concrete arch bridges are world-renowned not only because of their large spans, but also due to introduction and subsequent improvements in construction of reinforced concrete arches using the suspended cantilever technique (Radić et al., 2004). In fact, the Šibenik Bridge is the first concrete arch in the world which was erected entirely by the free cantilever method (Radić et al., 2008). The most famous among them is Krk Bridge, completed 35 years ago but still the largest spanning conventional reinforced concrete arch in the world.

Table 1 – Large span concrete arch bridges on the Adriatic coast

Bridge	Completed in	Span, L [m]	Rise, f [m]	f/L
Šibenik Bridge	1966	246	31	1/8
Pag Bridge	1968	193	28	1/7
Krk Bridge – larger arch	1980	390	60	1/6.5
Krk Bridge – smaller arch	1980	244	47	1/5.2
Maslenica Motorway Bridge	1997	200	65	1/3.1
Krka Bridge	2004	204	52	1/3.9
Cetina Bridge	2007	140	21.5	1/6.5

The first generation of concrete arch bridges (Figure 1) is related to the construction of the Adriatic National Road linking the coast and the mainland in 1970's and 1980's: Šibenik Bridge, Pag Bridge and Krk Bridge (two arches). The term second generation is used for three reinforced concrete arch bridges built in late 20th and early 21st century (Figure 2): Maslenica Bridge and Krka Bridge (Radić et al., 2010) on the Zagreb−Split motorway and Cetina Bridge on the connecting road between this motorway and the town Tril (Žderić et al., 2007).

Large span has not been the only challenge during construction of bridges on the Adriatic Sea. Aggressive maritime environment and high seismicity of Adriatic regions may have a negative impact on durability and reliability of the structures. The dangerous influences of marine environment on durability of the bridges are (Radić et al., 2003, Beslać et al., 2008):

Fig. 1 – The first generation of Adriatic large reinforced concrete arches: Šibenik Bridge (left), Pag Bridge (right) and Krk Bridge (bottom)

Fig. 2 – The second generation of Adriatic large reinforced concrete arches: Maslenica Bridge (left), Krka Bridge (middle) and Cetina Bridge (right)

- ingress of chlorides from the sea with rather high salinity between 3.5% and 3.8% of water mass ;
- local marine environment with the very frequent changes of very strong southern and northern (Bora) winds during winter months causing salt spray, depositing chlorides on all structural elements;
- summer temperature facilitating the chloride penetration and thus causing reinforcement corrosion, and
- temperatures below freezing point 10−15 times every winter.

Seismological map (Figure 3, left) represents design value of ground acceleration for return periods of 475 years. The values of long term expected peak ground acceleration are highest in the south coastal part (the Dinarides) and reach maximum values of up to 0.38 g. Wind load in some specific locations of Croatian coast, with its turbulent action together with local terrain shape, exceeds the maximum reference wind velocity of 35-40 m/s (Figure 3, right).

However, defects in structures of the first generation of Adriatic concrete bridges were found at an early age and hence cannot be blamed only on the aggressive environment, but may also be attributed to conceptual design and errors and negli-

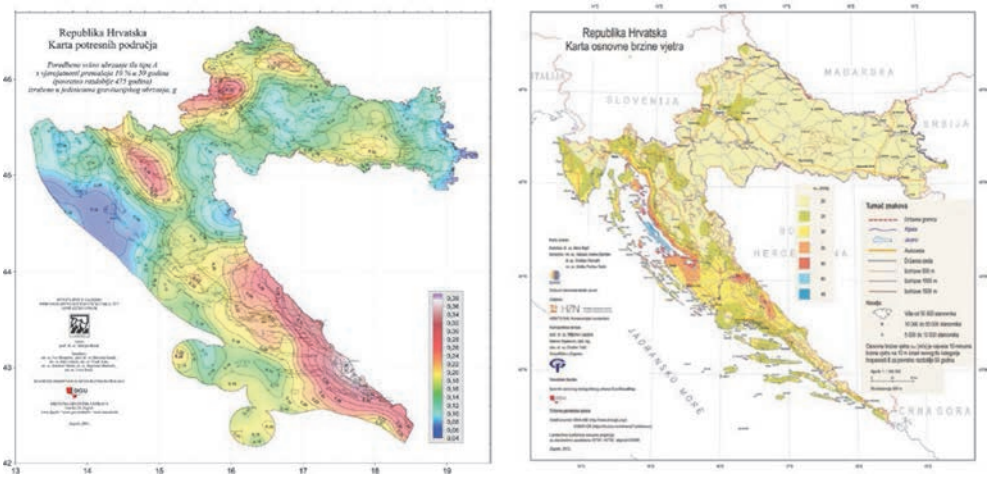


Fig. 3 – Seismological map (HRN EN 1998-1:2011/NA) (left) and basic wind velocity map (HRN EN 1991-1-4:2012/NA) (right) for Croatia

gence on site. Design errors can be summarized, as follows (Šavor et al., 2002, Radić et al., 2007):

- misinterpretation of the codes in designing very small concrete cover,
- extensive usage of precasting not supported by satisfactory constructional detailing,
- underestimated effect of creep and shrinkage,
- placing the precast prestressed girders of the superstructure directly on tops of piers with no fixed connection and no bearings, and
- inadequate drainage systems.

Extreme severity of the exposure conditions combined with the exceptionally small concrete cover, inadequate detailing, poor workmanship and maintenance led to the rapid structural degradation of the Pag Bridge (Šavor et al., 2008). Repairs on the Pag Bridge started already 10 years after the bridge completion, and finally the bridge had to be thoroughly reconstructed. The arch was repaired by removing the damaged concrete cover, grouting all visible cracks and placing an additional reinforcement mesh on all external arch surfaces covered by a fine 4-cm thick mortar layer. The prestressed concrete superstructure had to be replaced with the new continuous steel superstructure. Concrete columns were strengthened by encasing in steel casings 12 mm thick while leaving a 12-cm wide gap between the casing and the original column. This space was filled with fine aggregate low shrinkage concrete. To enhance the composite action shear connectors were placed on tops of the columns.

The application of concrete protection measures on the Krk Bridges started immediately after they were opened to traffic (Šavor et al., 2009). Almost all known

methods of repair and protection of reinforced concrete structures in marine environment were tried, but most of them failed and even accelerated the reinforcement corrosion, because 1-2 cm of very good concrete was replaced with very permeable mortar. By the end of the last century new improved flexible polymer coats for concrete protection from the ingress of chlorides appeared on the market and were applied to the exposed surfaces, which finally seem successful. The repair and protection of high bridge columns had to be done by removing 3-4 cm of concrete cover and by replacing it with 10 cm of high strength self-compacted concrete, micro-reinforced with 50 kg/m^3 of steel fibres.

Lessons learned from the first generation of arch bridges have been incorporated into design of more recent projects in order to assure durability and reliability of bridges in aggressive maritime environment with large seismicity.

The building of the new concrete arch Maslenica Bridge marked the beginning of construction of the new highway (Radić et al., 2010). The arch is of 200 m span with a rise of 65 m and was constructed by free cantilevering. The superstructure consists of eight simple-span precast prestressed girders made continuous and interconnected by concrete deck slab cast in situ. The number of structural joints has been reduced to a minimum, with expansion joints provided at the abutments only. The superstructure is supported by longitudinally movable bearings at both abutments and at the piers nearest to the abutments and by fixed bearings at the two piers near the arch crown. All other piers are fixed to the superstructure. Special measures have been taken to provide for the proper functioning of the bridge, situated in an aggressive maritime environment, until the end of its service life set at hundred years. Structural dimensions have been chosen adequately, in general much larger than on any previously built arch bridge on the Croatian coast of the Adriatic Sea. With the help of concrete materials engineers a low permeability concrete has been designed, with water-cement ratio w/c less than 0.40. For all bridge structural elements the minimum concrete cover of 5.0 cm was utilized, with the exception of the arch foundations nearest to the sea, where the minimum concrete cover was increased to 10.0 cm. Thirteen years after the Maslenica Bridge was opened to traffic, in October 2010, visual inspection and chloride content measurements were carried out on the bridge. The bridge is general in good condition and corrosion process in most structural members is still in the initiation stage; however considerable damages are observed on the columns at arch abutments: there are areas of exposed corroded reinforcement and at those locations concrete cover is less than 5 cm specified in the bridge design (Bleiziffer et al. 2011).

For the Krka Bridge a concrete arch of 204.0 m span and 52.0 m rise was selected, with the rise to span ratio of 0.25 (Radić et al., 2010). Instead of the usual prestressed concrete superstructure a composite superstructure was chosen to save weight thus reducing seismic forces and also to blend more harmonically into the

beautiful environment. An outstanding aesthetical impression was achieved, because cross beams connecting column tops were avoided.

In comparison to the first generation of arch bridges, bridge requirements have greatly changed: traffic loads are increased, regulations concerning seismic analysis are tightened and there is a new seismic categorization, which implies much larger earthquake forces (Figure 3, left).

A new approach to the problem of durability was also important for the structural design. The designers avoided wider use of prefabricated elements, concrete covers were increased up to three times, compared to Krk Bridges, and the forms of concrete elements were simplified in order to increase precision in production. It is not irrelevant to mention the new relations between prices of materials and construction works, which also affected the design philosophy, as well as newly developed scaffolding and formwork systems. The bridges are equipped with a range of sensors for long-term control of stresses, strains and corrosion progress with the intention to closely monitor both structural performance and durability related performance in order to facilitate the future maintenance activities by triggering timely adjustments and interventions (Rak et al., 2006, Šimunić et al., 2009).

3. Effects of corrosion on mechanical behaviour of RC structures

In-service performance of the first generation of the Adriatic arch bridges clearly indicates chloride-induced corrosion as the major cause of deterioration of RC structures in maritime environment. Corrosion process is recognized in reinforced concrete structures by brown patches of rust that emerge on the concrete surface and/or cracked concrete cover. Corrosion of steel in concrete is an electrochemical process which includes dissolution of iron and forming of corrosion products. Direct consequences of this process are: (i) reduction of the reinforcement bar area and ductility and (ii) volume expansion of corrosion products (fib, 2000; Cairns et al., 2005; Du et al., 2005a, 2005b). Reduction of the reinforcement bar area leads to decreased shear and moment capacities as well as decreased stiffness of the structure. A change in rebar ductility directly influences the stiffness of the structure, the possibility for force and moment redistribution, and limits the load-carrying capacity of a statically indeterminate structure (Rodriguez et al., 1997, Zandi Hanjari, 2010).

Pitting corrosion can significantly reduce cross section what causes reducing of load-bearing capacity and fatigue strength and under special circumstances may lead to brittle fracture of pre-stressing tendons made of high-strength steel (Kušter Marić, 2013).

Volume expansion of rust around reinforcement bar may cause the surrounding concrete to crack and spall off, which decreases the concrete cross-section and concrete cover. On the compressive side of a concrete element, spalling of the cover decreases the internal lever arm, which in turn decreases the bending moment. Furthermore, reduced confinement influences the interaction between the reinforcement and the concrete, which affects the anchorage capacity (Zandi Hanjari, 2010).

Cracked concrete surrounding corroded reinforcements and stirrups influences the anchorage and shear capacity of a beam. If the concrete in this region has been cracked by corrosion, it has reached its maximum tensile strength. Thus, any further tensile stress induced by mechanical loading contributes to opening larger cracks (Zandi Hanjari, 2010).

Cracked concrete not only affects actual shear and anchorage capacities, but also reduces the load-carrying capacity of a structure in the long-term by giving less protection to reinforcement and allowing an aggressive environment direct access to the reinforcement. Moreover, the cracks, depending on their direction, may also change the stiffness, and thereby altering the force distribution in the structure (Zandi Hanjari, 2010).

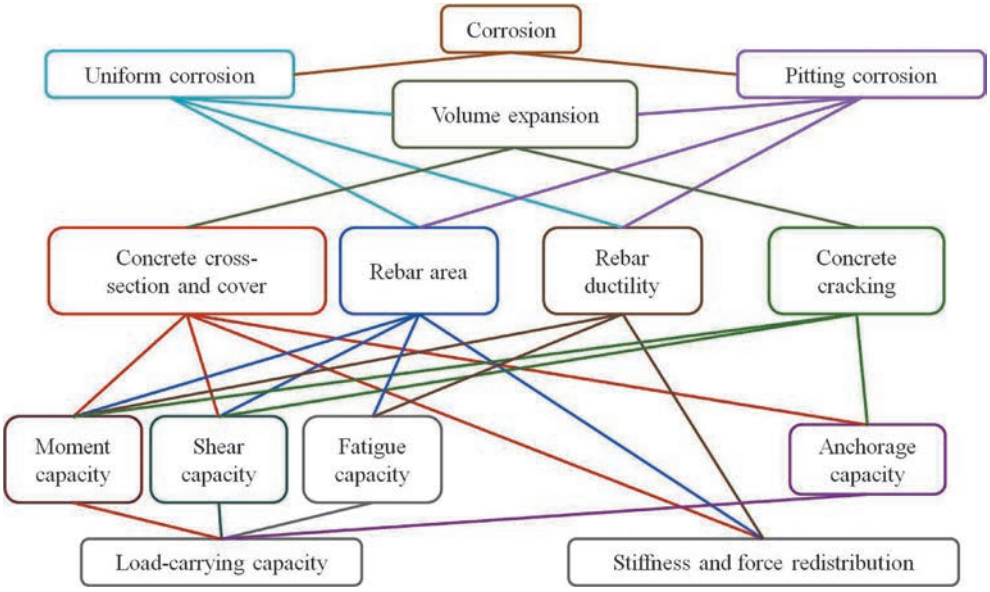


Fig. 4 – Effects of corrosion on mechanical behaviour of RC structures [Hanjari 2010]

Reduced cross-section of concrete column, as a result of cover spalling, may change slenderness and cause buckling of a concrete column (Rodriguez et al., 1997, Zandi Hanjari, 2010).

4. Chemo-hygro-thermo-mechanical model for concrete

Repairs and reconstruction works on the Adriatic arch bridges have been not only expensive, but technically demanding tasks and very difficult to perform. Therefore, to ensure efficient bridge management it is necessary to predict durability of new or already damaged reinforced concrete structures which requires a numerical model able to realistically simulate degradation processes and its effects on the structural safety.

As chloride-induced corrosion is the major cause of deterioration of reinforced concrete structure in maritime environment, deterioration prediction models during corrosion process on structural members is essential for efficient bridge management. A coupled 3D chemo-hygro-thermo-mechanical model for transient analysis of corrosion processes before and after depassivation of steel reinforcement in concrete has been developed and implemented into the 3D FE code for this purpose.

To estimate reduction of the reinforcement cross-section and to predict the volume increase of the corrosion product it is necessary to calculate the corrosion rate what requires modelling of the following physical, electrochemical and mechanical processes (Ožbolt et al., 2010):

- transport of capillary water, heat, oxygen and chloride through the concrete cover;
- immobilization of chloride in the concrete;
- cathodic and anodic polarization;
- transport of OH- ions through electrolyte in concrete pores;
- distribution of electrical potential and current density;
- transport of corrosion products in concrete and cracks;
- concrete cracking due to mechanical and non-mechanical actions.

$$\frac{\partial \theta_w}{\partial t} = \nabla \cdot \left[D_w(\theta_w) \nabla \theta_w \right] \tag{1}$$

where θ_w is volume fraction of pore water (m^3 of water / m^3 of concrete) and $D_w(\theta_w)$ is capillary water diffusion coefficient (m^2/s) described as a strongly non-linear function of moisture content (Leech et al., 2003). Transport of chloride ions through a non-saturated concrete occurs as a result of convection, diffusion and physically and chemically binding by cement hydration product (Bear, Bachmat, 1991):

$$\theta_w \frac{\partial C_c}{\partial t} = \nabla \cdot \left[\theta_w D_c(\theta_w, T) \nabla C_c \right] + D_w(\theta_w) \nabla \theta_w \nabla C_c - \frac{\partial C_{cb}}{\partial t} \tag{2a}$$

$$\frac{\partial C_{cb}}{\partial t} = k_r \left(\alpha C_c - C_{cb} \right) \tag{2b}$$

where C_c is concentration of free chloride dissolved in pore water (kg_{Cl}/m^3 pore solution), $D_c(\theta_w, T)$ is the effective chloride diffusion coefficient (m^2/s) expressed as a function of water content θ_w and concrete temperature T, C_{cb} is concentration of bound chloride (kg_{Cl}/m^3 of concrete), k_r is binding rate coefficient, $\alpha = 0.7$ is constant (Saetta et al., 1993).

Assuming that oxygen does not participate in any chemical reaction before depassivation of steel, transport of oxygen through concrete is considered as a convective diffusion problem (Bear, Bachmat, 1991):

$$, _w \frac{\partial C_o}{\partial t} = \nabla \cdot \left[\theta_w D_o(\theta_w) \nabla C_o \right] + D_w(\theta_w) \nabla \theta_w \nabla C_o \tag{3}$$

$$\lambda \Delta T + W(T) - c\rho \frac{\partial T}{\partial t} = 0 \tag{4}$$

where λ is thermal conductivity (W/(m K)), c is heat capacity per unit mass of concrete (J/(K kg)), ρ is mass density of concrete (kg/m^3) and W is internal source of heating (W/m^3).

The corrosion of steel is activated with the depassivation of the steel reinforcement in concrete. The non-mechanical processes important for the propagation stage of steel corrosion in concrete are:

- mass sinks of oxygen at steel surface due to cathodic and anodic reaction,
- the flow of electric current through pore solution and
- the cathodic and anodic potential.

The oxygen consumption at the cathodic and anodic surfaces is a result of the following reactions:

a) reaction of dissolved oxygen in the pore water with the electrons on the cathode

$$2H_2O + O_2 + 4e^- \rightarrow 4OH^- \tag{5}$$

b) transport of hydroxyl ions to the anode, where corrosion products forms

$$Fe^{2+} + 2OH^- \rightarrow Fe(OH)_2 \tag{6}$$

$$4Fe(OH)_2 + O_2 + 2H_2O \rightarrow 4Fe(OH)_3 \tag{7}$$

can be calculated as:

$$D_o(S_w, p_{con}) \frac{\partial C_o}{\partial n}\bigg|_{cathode} = -k_c i_c \qquad k_c = 8.29 \times 10^{-8} \frac{\text{kg}}{\text{C}} \tag{8a}$$

$$D_o(S_w, p_{con}) \frac{\partial C_o}{\partial n}\bigg|_{anode} = -k_a i_a \qquad k_a = 4.14 \times 10^{-8} \frac{\text{kg}}{\text{C}} \tag{8b}$$

where n is outward normal to the steel bar surface and i_c and i_a are cathodic and anodic current density (A/m^2), respectively. The constants k_c and k_a are calculated using the stoichiometry of chemical reactions (5-7) and Faraday's law.

According to Butler – Volmer kinetics, in the present model kinetics of reaction at the cathodic and anodic surface can be estimated from:

$$i_c = i_{0c} \frac{C_o}{C_{ob}} e^{2.3(\Phi_{0c} - \Phi)/\beta_c} \qquad i_a = i_{0a} e^{2.3(\Phi - \Phi_{0a})/\beta_a} \tag{9}$$

where C_{ob} is oxygen concentration at surface of concrete element exposed to sea-water (kg/m^3), Φ is electric potential in pore solution near reinforcement surface (V), i_{0c} and i_{0a} are the exchange current density of the cathodic and anodic reaction (A/m^2), Φ_{0c} and Φ_{0a} are the cathodic and anodic equilibrium potential (V), β_c and β_a are the Tafel slope for cathodic and anodic reaction (V/dec), respectively.

The electric current through the electrolyte is a result of motion of charged particles and, if the electrical neutrality of the system and the uniform ions concentration are assumed, can be written as:

$$\mathrm{i} = -\sigma(S_{w,} p_{con}) \nabla \Phi \tag{10}$$

where σ is electrical conductivity of concrete.

The equation of electrical charge conservation, if the electrical neutrality is accounted for and the electrical conductivity of concrete is assumed as uniformly distributed, reads:

$$\nabla^2 \Phi = 0 \tag{11}$$

More details related to the strong and weak formulations of the processes before and after depassivation of reinforcement can be found in Ožbolt et al., 2010, 2011.

Rate of rust production J_r (kg/m^{2}s) and mass of hydrated red rust per related surface (A_r) of rebar m_r (kg), respectively, are calculated as:

$$J_r = 5.536 \times 10^{-7} i_a$$
$$m_r = J_r \Delta t A_r \tag{12}$$

where Δt is time interval in which the corrosion is taking place and A_r is the corresponding surface of the steel reinforcement. The coefficient of proportionality between the anodic current density i_r and rate of rust production J_r is calculated using the stoichiometry of chemical reactions and Faraday's law (Ožbolt et al., 2011, 2012a, 2012b).

Recent experimental investigations (Wong et al., 2010, Fischer, 2012) have shown that the penetration of corrosion products into the pores and their relatively large ingress through the radial cracks, generated around the bar, has a significant effect on the development of corrosion induced damage. The influence can be summarized as:

- the distribution of rust and radial pressure over the anodic surface is not uniform and
- damage due to expansion of products is less pronounced.

The distribution of corrosion product (red rust) R (kg/m^3 of pore solution) into the pores and through the cracks in concrete has been mathematically formulated as a convective diffusion problem:

$$\theta_w \frac{\partial R}{\partial t} = \nabla \cdot \left[\theta_w D_r \nabla R \right] + D_w(\theta_w) \nabla \theta_w \nabla R \tag{13}$$

in which D_r is the diffusion coefficient (m^2/s) of corrosion product. It is important to note that the Eq. (13) does not directly describe the transport of the red rust, but rather the distribution of the rust formed in the concrete pores and cracks as a consequence of soluble species (which can dissolve in concrete pore solution and subsequently migrate in pores and cracks) reacting with oxygen in the pore water (Wong et al., 2010).

Currently, the above mentioned process of transport of corrosion products is considered more in a qualitative sense. Detailed experimental investigations needed for the quantitative calibration of the model are currently not available. Therefore, further development of the model is planned in the context of experimentally measuring the transported amount of corrosion products into the cracks and then using this data for the adjustment of the model parameters.

The microplane model for concrete with relaxed kinematic constraints (Ožbolt et al., 2001) is applied in the mechanical part of the model. In the finite element anal-

ysis cracks are treated in a smeared way, i.e. smeared crack approach is employed. To assure the objectivity of the results with respect to the size of the finite elements, the crack band method is used (Bažant, Oh, 1983).

The governing equation for the mechanical behaviour of a continuous body in the case of static loading condition reads:

$$\nabla\left[D_m\left(u,\theta_w,T\right)\nabla u\right]+\rho b=0 \tag{14}$$

in which D_m is material stiffness tensor, ρb is specific volume load and u is displacement field. In the mechanical part of the model the total strain tensor is decomposed into mechanical strain, thermal strain, hygro strain (swelling–shrinking) and strain due to expansion of corrosion products.

One-dimensional corrosion contact elements are employed in the model to account for the inelastic strains due to the expansion of corrosion products. They are placed radially around the bar surface and their main function is to simulate the contact between reinforcement and the surrounding concrete. These contact elements can take up only shear forces in direction parallel to reinforcement axis and compressive forces perpendicular to the adjacent surface of the reinforcement. The inelastic radial expansion due to corrosion Δl_r is calculated as:

$$\Delta l_r=\frac{m_r}{A_r}\left(\frac{1}{\rho_r}-\frac{0.523}{\rho_s}\right) \tag{15}$$

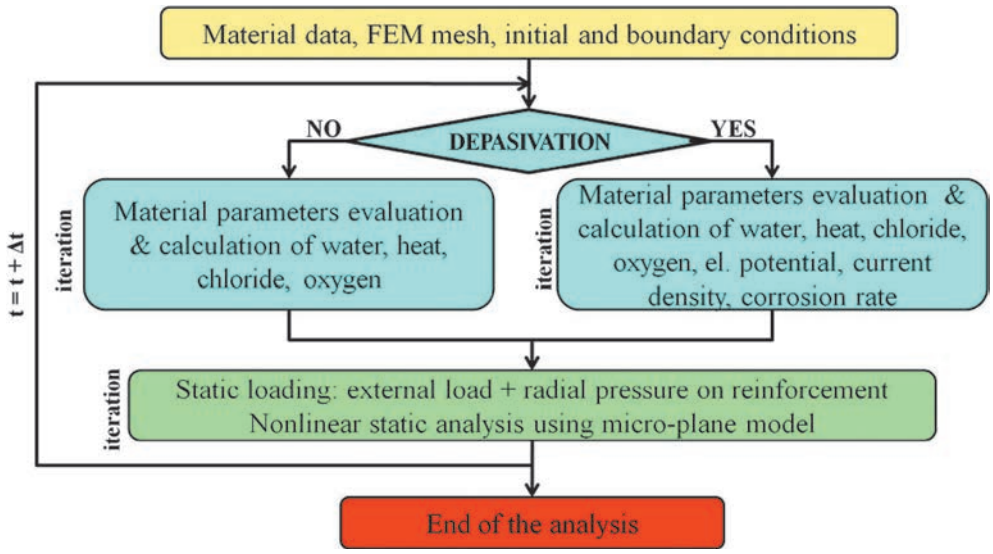


Fig. 5 – Numerical algorithm of the 3D chemo-hygro-thermo mechanical model

where $\rho_r = 1.96 \times 10^3$ (kg/m^3) and $\rho_s = 7.89 \times 10^3$ (kg/m^3) are densities of rust and steel, respectively, 0.523 is the ratio between the mass of steel (m_s) and the corresponding mass of rust (m_r) over the related surface of reinforcement A_r that corresponds to the contact element.

The stiffness of the rust layer is assumed to be $E_r = 100$ MPa. In the model it is represented by the axial stiffness of the corrosion contact elements. The shear resistance of the contact elements, defined by the bond-slip relationship, is used to model the bond between deformed steel reinforcement and concrete.

Algorithm of the model is shown on figure 5. The numerical analysis starts after the geometry, initial and boundary conditions have been defined and finite element mesh generated. After material parameters evaluation, transport of water, heat, free chloride and bound chloride and oxygen are calculated. If depassivation condition is satisfied, electric potential and current density will be also calculated. The numerical analysis is incremental. In each time or load step partial differential equations of non-mechanical and mechanical part of the model (equation of equilibrium) are solved simultaneously. When solving non-mechanical part of the problem, mechanical parameters are constant, equal to those of the previous time step, and conversely (Kušter Marić, 2013).

4. Application of the numerical model

The applications of the 3D chemo-hygro-thermo-mechanical model are illustrated on three numerical examples. The aim of first numerical example is to demonstrate the influence of cracks in concrete on transport processes and on depassivation time (Ožbolt et al., 2010).

Simply supported reinforced concrete slab was first damaged by external load and subsequently exposed to aggressive influence of seawater at the bottom side (Figure 6).

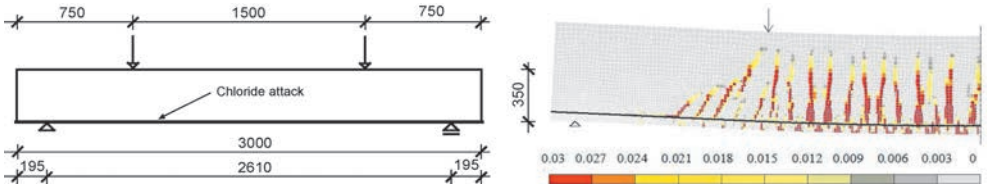


Fig. 6 – Geometry of analyzed reinforced concrete slab with concrete cover of 30 mm (left) and FE discretization and distribution of cracks with red zones as maximal principal strains in the slab (right)

The distribution of free chloride, in time sequences between 1 hour and 10 years, is shown in Figure 7 (left). The analysis predicts depassivation time of reinforcement (concentration of free chloride: 7.0 kg/m^3 of pore water) in the cracked zone immediately after the crack formation. Contrary to this, for the uncracked part of the slab depassivation time is not reached even after 10 years. Comparing the distribution of the chlorides at different times (Figure 7, right), it can be seen that the free chlorides penetrate in the region between the cracks (horizontal direction). Therefore, there is a slight decrease of their concentration in the crack, i.e. with increase of time chlorides tend to be smeared-out into the horizontal direction. A similar result was observed in experiments (Marsavina et al., 2009).

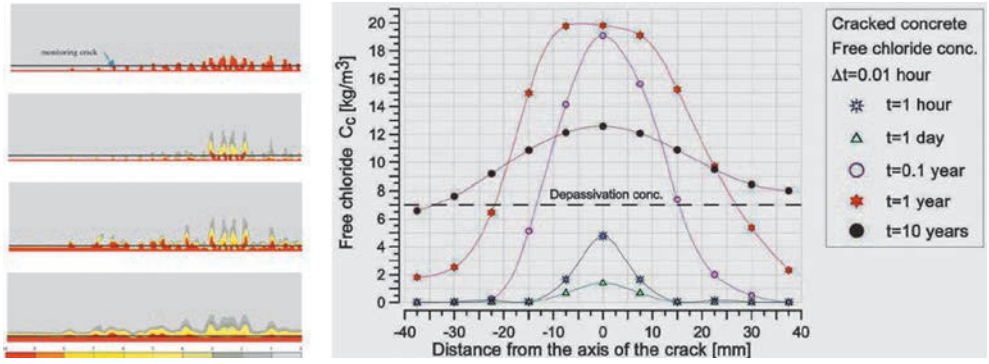


Fig. 7 – Distribution of free chlorides after 1 hour, 1 day, 1 year and 10 years (left) and distribution of free chlorides at the level of reinforcement, left and right of the crack (right)

In second numerical example processes after depasivation of reinforcement are analysed (Ožbolt et al., 2011). The study is performed for a concrete beam, cracked under axial tensile loading and exposed to splash zone. The aim of the study was to demonstrate the influence of concrete quality, water saturation and cracks in concrete on corrosion current density of a macro cell (Figure 8).

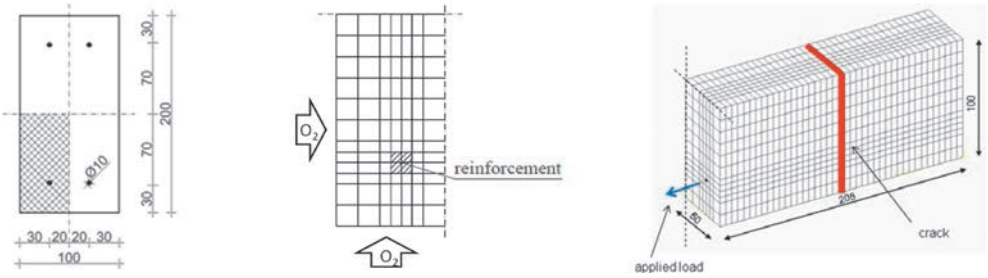


Fig. 8 – Geometry (all in mm), mechanical and non-mechanical loading, boundary conditions and finite element discretization

The 3D distributions of oxygen, electric potential and corrosion current density on both sides of the vertical section of the beam through the steel reinforcement bar, for un-cracked and cracked concrete (w/c = 0.4, S= 45%) after 10 minutes of corrosion process are shown in Figure 9. The distribution of electric potential and current density indicates the pitting corrosion in the anode-cathode transition zone, which is especially dangerous because of rapid reduction of the reinforcement cross-section area and strong degradation of ductility.

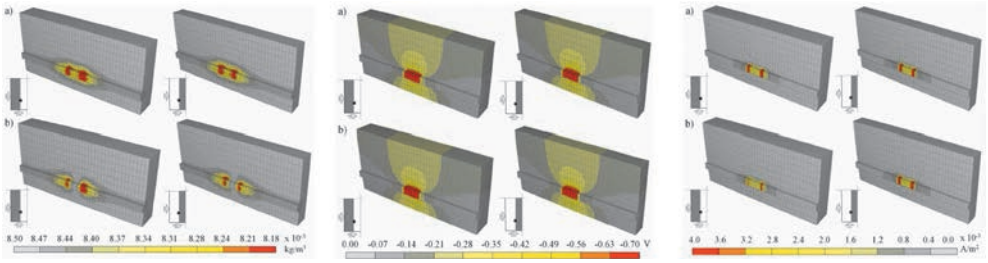


Fig. 9 – 3D distribution of oxygen (left), electric potential (middle) and current density (right) for good quality concrete at saturation of 45% after 10 minutes of corrosion process for: a) un-cracked and b) cracked concrete

Maximal consumption of oxygen is predicted in the transition zone between anode and cathode (Figure 10). The oxygen distribution in cracked and un-cracked concrete is very similar. The only difference is noticed in the vertical row of elements representing crack. There is almost no consumption of oxygen because of a continuous oxygen supply in the crack. The consumption of oxygen is much faster in a good quality concrete (w/c=0,4) than in poor quality concrete (w/c=0,7), since water and oxygen diffusivities increase with increase of the water to cement ratio. Furthermore, the consumption of oxygen increases with increase of water saturation up to critical saturation. In the good quality concrete the critical water saturation, at which the concentration of oxygen at cathodic site reaches a small positive value and simulation is still stable, is approximately 70%. Similar results can be observed for poor quality concrete; however, the oxygen consumption is much slower and the critical water saturation is 85%. In all analyzed cases the highest

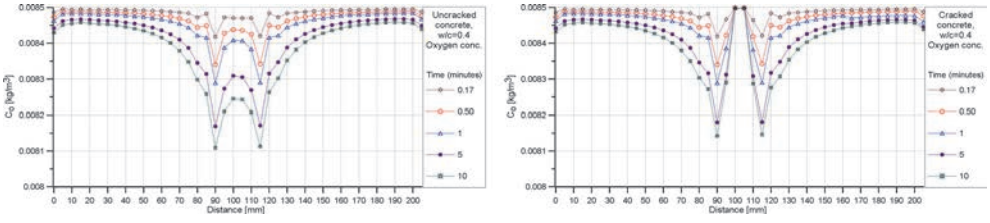


Fig. 10 – Oxygen concentration at the level of reinforcement in un-cracked (left) and cracked (right) good quality concrete

corrosion rate is obtained in the anode-cathode transition zone. The corrosion rate is higher in poor than in good quality concrete and cracking does not have significant influence on the maximal current density (Figure 11, left). Numerical results of current density for un-cracked, good quality concrete are in good agreement with the experimental results (Lopez, Gonzales, 1993). The used mathematical models for corrosion kinetics and transport of ions simulate the process of reinforcement corrosion in unsaturated concrete for water saturation in the range from 35% up to the critical value (depending on concrete properties) in which corrosion is controlled by both, reactive and transport process. The diffusion-controlled corrosion of steel in concrete takes place in the range of water saturation greater than the critical saturation (not considered in the present research). In such a case the reduction of oxygen is faster than its supply.

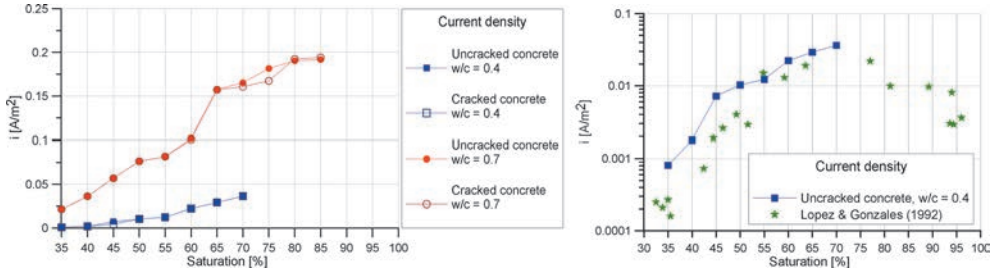


Fig. 11 – Relation between saturation and current density measured at the anode-cathode transition zone: for all calculated cases (left) and comparison with experiments (right)

It is known that corrosion current density of 0,05 A/m^2 to 1,00 A/m^2 is equivalent to the reduction of reinforcement diameter for approximately 0,1 to 2,0 mm/year, respectively (Martín – Pérez, 1999, Bäßler, Burkert, 2003). These limits are approximately valid for concrete in the splash zone. The corrosion current densities predicted in the numerical simulations are within the above limits. This confirms that the prediction is realistic and that the splash zone is one of the most critical areas for corrosion of steel in concrete.

The aim of the third numerical example is to demonstrate influence of rust transport into the concrete cracks and pores on concrete cracking as well as influence of corrosion on bond resistance (Ožbolt et al., 2012a). Two beam-end specimens with a cross-section of 200 x 200 mm and four bars placed in the corners are chosen. The diameter of the reinforcement bars are 12 mm and 16 mm with a concrete cover of 20 mm and 35 mm in the first and the second specimen, respectively (Figure 12). To reduce computational time only processes after depassivation are analyzed and only 1/2 of the beam cross section is discretized by eight-node solid 3D finite elements. Several anodic and cathodic regions – macro cells are assumed along the reinforcement length embedded into concrete (Figure 12).

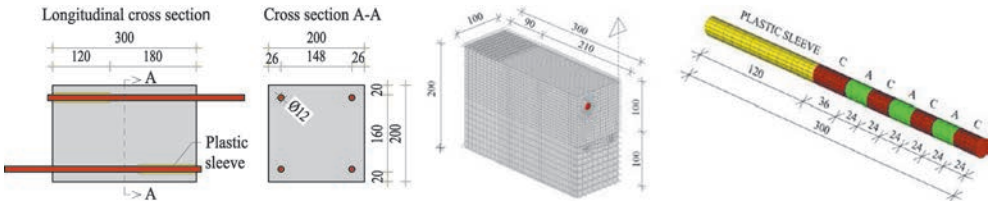


Fig. 12 – The 1st specimen: cross-section (left), finite element discretization of the model (middle) and assumed macro cells on the corroded steel reinforcement (right)

Electric potential, current density and distribution of oxygen are calculated for each analysed case. To study the influence of expansion of corrosion products on concrete cracking, two solutions are analyzed for each specimen: with and without transport of rust through cracks. Cracking of concrete cover is less pronounced if the transport of rust through cracks is accounted for. Numerical results show very good agreement with experimental tests (Figure 13) in which, for assumed environmental conditions, approximately 50 % of corrosion products were transported into cracks (Fischer, 2012). As numerical analysis shows, this significantly reduces corrosion induced cracking of concrete cover and contributes to higher pull-out capacity of corroded steel reinforcement. Similar crack patterns were observed in the experimental tests in Fischer, 2012 (Figure 13), however, in the tests the corrosion was accelerated with the rate that is approximately 20 times higher than natural corrosion rate. Therefore development of corrosion induced cracking in time cannot be directly compared with numerical prediction in which the natural corrosion conditions are assumed (aggressive, splash zone).

To study the influence of corrosion induced cracks on the bond resistance, steel reinforcement bar is pulled out from the first and second specimen at t =0 (reference), 1, 3, 5 and 7 years. With the increase of corrosion rate the pull-out capacity decreases. The decrease is higher if no transport of corrosion products is assumed.

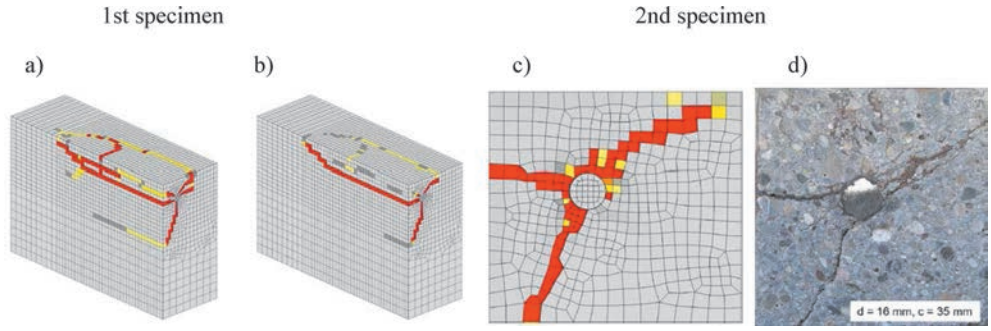


Fig. 13 – Cracks 7 years after depassivation: (1) in the 1st specimen: without (a) and with rust transport (b); and (2) in the middle of the 2nd specimen: (c) numerical study, (d) experimental results by Fischer, 2012

There is a nice agreement between experiments and calculations (Figure 14). It should be noted that numerical results with transport of rust should be actually compared with test data. The reduction of the pull-out capacity and the influence of the transport of rust through cracks are more significant in the case of the specimen with larger bar diameter (ϕ16/35 mm) (Figure 14) (Ožbolt et al., 2012a).

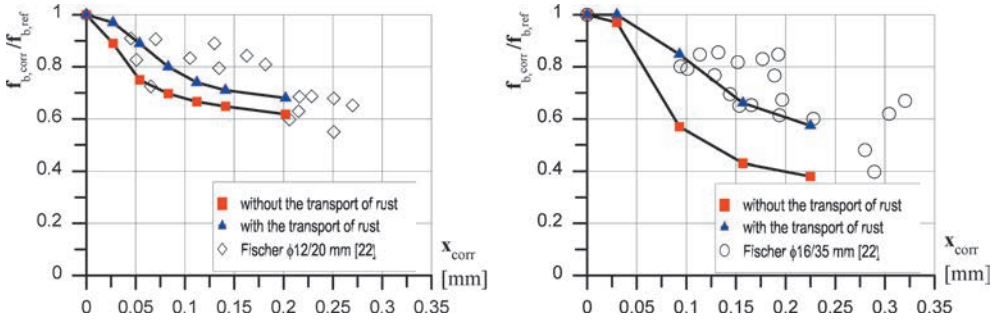


Fig. 14 – Relation between loss of steel reinforcement due to corrosion in mm and pull-out capacity for (a) the first specimen (ϕ12/20 mm) and (b) the second specimen (ϕ16/35 mm) (Ožbolt et al., 2012a)

6. Conclusion

The older Adriatic large concrete arch bridges suffered greatly over decades of service, due to combination of aggressive maritime conditions (high winds, chloride ingress, freeze/thaw and thermal action) and inadequate attention to durability issues. Serious deterioration of structural members with reinforcement corrosion being the major issue, led to many complex and expensive repairs.

In order to develop consistent maintenance policy that would provide efficient and effective management of bridges in maritime environment, it is necessary to determine remaining service life of the structures. In this purpose the fully coupled 3D chemo-hygro-thermo-mechanical model for transient analysis of corrosion processes before and after depassivation of steel reinforcement in concrete is developed and implemented into 3D FE code. The applications of the model are illustrated on three numerical examples. The numerical results are in good agreement with the available experimental data leading to the conclusion that the model is able to realistically predict service life of reinforced concrete structures exposed to chlorides and mechanical damage results.

In service performance of the older Adriatic large concrete arch bridges shown and numerical analysis confirmed the conclusions listed below.

Cracks in concrete cover significantly reduce depassivation time of reinforcement bar. Defects in structures of the first generation of Adriatic concrete bridges were found at an early age and they are consequence of mistakes in conceptual design as well as errors and negligence on site. Therefore, it is not surprising that the problems related to corrosion of steel in concrete occurred few years after the opening of the bridge. The numerical analysis predicts depassivation time of reinforcement in the cracked zone immediately after the crack formation. Contrary to this, for the uncracked part of the slab depassivation time is not reached even after 10 years.

Numerical analysis shows that corrosion rate is much higher in poor than in good quality concrete. The same issue was in application of concrete protection measures on the Krk bridges, because 1-2 cm of very good concrete was replaced with very permeable mortar what accelerated the reinforcement corrosion.

The maximal values of corrosion rate in the second numerical example, achieved at critical water saturation, are within the limits for concrete in splash zone: 0,05 A/m^2 to 1,00 A/m^2, what is equivalent to the reduction of reinforcement diameter for approximately 0,1 to 2,0 mm/year.

Corrosion process is recognized in reinforced concrete structures by brown patches of rust that emerge on the concrete surface and/or cracked concrete cover. The third numerical example shows that corrosion induced cracks strongly reduces pull-out capacity of reinforcement. Transport of rust through cracks reduces unfavourable effects of corrosion induced cracking of concrete cover, but it must be included in numerical model in order to more precisely simulate corrosion effects on reinforced concrete structures.

Further development of the model is in progress.

References

[1] Bäßler, R., Burkert, A. (2003). GPM-portable equipment for determination of corrosion stage of concrete structures – laboratory and on-site experiences. In: Proceedings of International Symposium Non-Destructive Testing in Civil Engineering (NDT-CE), Berlin.

[2] Bažant, Z.P., Oh, B.H. (1983). Crack band theory for fracture of concrete. Mater Struct 16(93) 1983, pp. 155–177.

[3] Bear, J., Bachmat, Y. (1991). Introduction to modeling of transport phenomena in porous media. Boston: Kluwer Academic Publishers.

[4] Beslać, J., Tkalčić, D., Štemberga, K. (2008). Difficulties and Successes in the Maintenance of Krk Bridge. In: Radić, J., Chen, B.C. (eds), Long Arch Bridges, pp. 197-205. Zagreb: Secon HDGK.

[5] Bleiziffer, J., Mavar, K., Ille, M., Škarić Palić, S., Dimic Vukovic, S., Balagija, A. (2011). Condition Assessment of the Maslenica Highway Bridge. In: Radić, J., Chen, B.C. (eds), Sustainable Arch Bridges, pp. 309-316. Zagreb: Secon HDGK.

[6] Cairns, J., Plizzari, G. A., Du, Y., Law, D. W., Franzoni, C. (2005) Mechanical properties of corrosion-damaged reinforcement. ACI Materials Journal, 102 (4) 2005, pp. 256-264.

[7] Du, Y. G., Clark, L. A., Chan, A. H. C. (2005a) Effect of corrosion on ductility of reinforcing bars. Magazine of Concrete Research, Vol. 57 (7) 2005, pp. 407-419.

[8] Du, Y. G.; Clark, L. A. & Chan, A. H. C. (2005b) Residual capacity of corroded reinforcing bars. Magazine of Concrete Research, Vol. 57 (3) 2005, pp. 135-147.

[9] fib (2000) Bond of Reinforcement in Concrete – State-of-art report. Prepared by fib Task Group Bond Models. fib Bulletin 10. Lausanne: International Federation for Structural Concrete.

[10] Fischer C. Influences of reinforcement corrosion on bond between steel and concrete. PhD thesis. Stuttgart,2012. (in German)

[11] HRN EN 1991-1-4:2012/NA 2012. Eurocode 1: Actions on structures — General actions — Part 1-4: Wind actions – National Annex. Zagreb: Croatian Standards Institute. (in Croatian).

[12] HRN EN 1998-1:2011/NA 2011. Eurocode 8: Design of structures for earthquake resistance – Part 1: General rules, seismic actions and rules for buildings – National Annex. Zagreb: Croatian Standards Institute. (in Croatian).

[13] Leech, C., Lockington, D., Dux P. (2003). Unsaturated diffusivity functions for concrete derived from NMR images. Mater Struct 36, 2003, pp. 413-418.

[14] Lopez, J.A., Gonzales, K. (1993) Influence of the degree of pore saturation on the resistivity of concrete and the corrosion rate of steel reinforcement, Cement and Concrete Research, 23, 1993, pp. 368–376.

[15] Kušter Marić, M. (2013). Service life prediction of reinforced concrete bridges exposed to chlorides. PhD thesis, University of Zagreb, Croatia. (in Croatian).

[16] Marsavina L., Audenaert K., De Schutter G., Faur N. And Marsavina D. (2008). Experimental and numerical determination of the chloride penetration in cracked concrete. Construction and Building Materials 23, 2008, pp. 264-274.

[17] Martín – Pérez, B. (1999). Service life modelling of RC highway structures exposed to chlorides. PhD thesis, University of Toronto, Canada.

[18] Ožbolt, J., Li, Y., Kožar, I. (2001). Microplane model for concrete with relaxed kinematic constraint. Int J Solids Struct, 38(16) 2001, pp. 2683–2711.

[19] Ožbolt, J., Balabanić, G., Periškić, G., Kušter, M. (2010). Modelling the effect of damage on transport processes in concrete. Constr Build Mater 24(9) 2010, pp. 1638–1648.

[20] Ožbolt, J., Balabanić, G., Kušter, M. (2011). 3D Numerical modelling of steel corrosion in concrete structures. Corros Sci 53(12) 2011, pp. 4166–4177.

[21] Ožbolt, J., Oršanić, F., Kušter, M., Balabanić, G. (2012a). Modelling bond resistance of corroded reinforcement. In: Cairns, J.W., Plizzari, G. (eds). General aspects of bond: Bond in Concrete, pp. 437–444. Publisher Creations.

[22] Ožbolt, J., Oršanić, F., Balabanić, G., Kušter, M. (2012b). Modeling damage in concrete caused by corrosion of reinforcement: coupled 3D FE model. Int J Fract 178(1-2): 2012, pp. 233–244.

[23] Radić, J., Šavor, Z., Puž, G. (2003). Extreme Wind and Salt Influences on Adriatic Bridges, Structural Engineering International. 13(4) 2003, pp. 242-245.

[24] Radić, J., Žderić, Ž., Puž, G. (2004). Construction methods for reinforced concrete arch bridges. In Katalinic, B. (ed.) DAAAM International Scientific Book, pp. 519-536. Vienna: DAAAM International.

[25] Radić, J., Bleiziffer, J., Kušter, M. (2007). Trends and developments in *bridge* and asset *management systems,* In Celikag, M. (ed.), Proceedings of the 11th International Conference on Inspection, Appraisal, Repairs and Maintenance of Structures, pp. 43-53. Singapore: CI-Premier Pte Ltd.

[26] Radić, J., Šavor, Z., Bleiziffer, J., Kalafatić, I. (2008). Šibenik Bridge – Design, Construction and Assessment of Present Condition. In: Radić, J., Chen, B.C. (eds), Long Arch Bridges, pp. 207-216. Zagreb: Secon HDGK.

[27] Radić, J., Šavor, Z. & Mandić, A. (2010). Two Notable Arch Bridges on The Croatian Adriatic Highway. Structural Engineering International. 20 (1) 2010, pp.36-40.

[28] Rak, M., Bjegović, D., Kapović, Z., Stipanović, I. & Damjanović, D. (2006). Durability monitoring system on the bridge over Krka river. In J. Radić (ed.) Bridges: 1137-1146. Zagreb: Secon HDGK.

[29] Rodriguez, J.; Ortega, L. M. & Casal, J. (1997) Load carrying capacity of concrete structures with corroded reinforcement. Construction and Building Materials, Vol. 11 (4) 1997, pp. 239-248.

[30] Saetta, A.V., Scotta, R.V., Vitaliani, R.V. (1993). Analysis of chloride diffusion into partially saturated concrete, ACI Mater. J. 90 (M47) 1993, pp. 441 – 451.

[31] Šavor, Z., Radić, J. & Puž, G. 2002. Maintenance and Repair of Large Concrete Arch Bridges. In Casas, J.R., Frangopol, D.M., Nowak, A.S. (eds) Proceedings IABMAS'02: Bridge Maintenance, Safety and Management. Pp. 1-8. Barcelona: International Center for Numerical Methods in Engineering.

[32] Šavor, Z., Mujkanović, N., Hrelja, G., Bleiziffer J. (2008). Reconstruction of the Pag Bridge. In: Radić, J., Chen, B.C. (eds), Long Arch Bridges, pp. 241-251. Zagreb: Secon HDGK.

[33] Šavor, Z., Šavor, M., Srbić, M. (2009). Krk bridge from inception to today. In: Radić, J., Chen, B.C. (eds), Construction of arch bridges, pp. 377-395. Zagreb: Secon HDGK.

[34] Šimunić, Z., Gašparac, I., Pavlović, B. (2009). Monitoring of Maslenica Bridge during construction. In Proceedings of IABSE Symposium Structures for the Future – The Search for Quality. Rio de Janeiro: IABSE.

[35] Wong, H.S., Zhao, Y.X., Karimi, A.R., Buenfeld, N.R., Jin, W.L. (2010). On the penetration of corrosion products from reinforcing steel into concrete due to chloride-induced corrosion. Corros Sci 52(7) 2010, pp. 2469–2480.

[36] Zandi Hanjari, K. (2010) Structural Behaviour of Deteriorated Concrete Structures. PhD thesis. Gothenburg, Sweden: Chalmers University of Technology – Department of Civil and Environmental Engineering – Division of Structural Engineering, Concrete Structures.

[37] Žderić, Ž., Runjić, A., Hrelja, G. (2007). Design and construction of Cetina River arch bridge. In Lourenco, P.B., Oliveira, D.V., Portela, A. (eds), Proceedings of the 5th Int. Conf.on Arch Bridges, pp. 745-750. Madeira: Multicomp Lda.

Professional paper

Mapping of Underwater Habitats Based on the Analysis of Backscatter Intensity of the Return Acoustic Signal

Boško Pribičević[*], Almin Đapo[*], Branko Kordić[*], Nino Pijanović[*]

University of Zagreb, Faculty of Geodesy

Mapping of underwater habitats, through a multidisciplinary approach, represents a fundamental prerequisite for adequate representation of physical and biological topologies with the objective of sustainable use and management of marine resources.

Increased demand for diverse geoinformation resulted with the development of modern technologies and the acquisition of new knowledge in the field of underwater acoustics as the basic method of modern hydrographic survey. In addition to 3D modeling of the seabed, modern multibeam sonars allow possibility of processing the propagation and backscatter data of the acoustic signal trough the water column. By integration with the classical methods of underwater habitat mapping, complete coverage of the seabed with reliable high resolution data is effectively achieved. The collected data represents the basis for further advanced interdisciplinary analysis in GIS environment.

The paper is theoretically based and methodologically appropriate representation of hydrographic surveying for the purpose of underwater habitat mapping in part of the southern coast of the Stari Grad bay on the island of Hvar in order to determine the state of underwater habitats of Posidonia seagrass.

The paper describes the theoretical and empirical basis of the acoustic signal backscatter whose specificity is used here as an identifier type of sea bottom that is of underwater habitats. Key steps in the creation of maps of underwater habitats are discussed theoretically and practically presented on the basis of the data collected. Appropriate map of underwater habitats containing three basic classes based on data sampling was made. Conclusions and

[*]bpribic@geof.hr , adapo@geof.hr, bkordic@geof.hr, nipijanovic@geof.hr

final observations which were derived through analysis of the results could show very useful in further study of the spatial distribution of underwater habitats.

Key words:

Mapping, habitats, multidisciplinary approach, underwater acoustics, backscatter

1. Introduction

The Adriatic Sea, with a significant role for the geographical identity of the country, represents one of the most valuable resources of Croatia. Many economic activities are related to the underwater ecosystem, however, at the same time they are creating a real danger for the destruction of the biodiversity. Human impact on the environment of the seabed has reached an unprecedented level. Human impact on the Adriatic through industrial operations, tourism, maritime transport and the practice of fishing is very little, or completely uncontrollable. Exploitation of undersea resources therefore directly depends on the preservation of long-term stability of the underwater ecosystem whose nature is very dynamic. Access to the data related to sea is vital for the marine industry, decision-making bodies and scientific research.

The rapidly increasing trend on environmental awareness has led to the general public less-known networks of protected habitats at the level of Europe or the world. NATURA 2000 and EMODnet represent the basis of the Europe's proactive approach on the protection of the underwater, and preservation of underwater endangered and native species through the mapping of the underwater habitats. Areas of the sea, covered by these networks in Croatia, are mainly related to the Croatian national parks, because of their biodiversity and fundamental role in the conservation of the species. The initiative for the formation of a network of this type is usually identified with the protected areas where human activities are significantly limited or even excluded in order to preserve biodiversity. This creates a misconception of the importance of these projects that are based on other principles. The objective of the management is sustainability, improving conservation of target species and habitats in a particular area, and in the case of the management of the areas of interest, the well-being of people who live in them is taken into the account. This is possible to achieve without prescribing essential constraints, but through the implementation of precautions by the people who are in touch with the nature and are sharing their living space with the targeted species.

Informatization of society is a global phenomenon. Practically there are no activities where information technology does not play a major role. The same have an

enormous impact in the scientific research. Therefore, a society that, in terms of informatization, does not keep pace with the developed countries, and especially with the surrounding regions and countries, lags behind in science. Geodesy and hydrography are also influenced by the development of new technologies, and many new research opportunities open up. The creation of systematically organized database of underwater is a unique goal for establishing the system of protection and control of the underwater. The same can be useful for a variety of scientific fields and is the foundation for the implementation of various activities or research in the area of interest.

Therefore, this paper describes a pilot project of the mapping of the underwater habitats by using the acoustic scattering data collected by remote sensing equipment of the new generation. In order to determine the state of the habitats of the sea grass Posidonia on the southern coast of the Stari Grad bay on the island of Hvar, a hydrographic survey was conducted using the modern multibeam sonar. On the basis of the conducted analysis of the acoustic backscatter signal in the area of the hydrographic survey, an exemplary map of underwater habitats containing three basic classes based on data sampling had been made. By analyzing the results, conclusions and final remarks were made that may be useful in further study of the spatial distribution of underwater habitats.

2. Mapping of the underwater (seabed) habitats

Mapping of the underwater (seabed) habitats is the current process in the world that is gaining more and more interest every day. The term “underwater habitat” is originally associated with the place where underwater animals or algae live (single specie). The original term can be extended by “merging” multiple species in one living community. The use of the term “habitat”, in terms of mapping, involves the physical and natural factors that support the living conditions of some of the communities. In this way the algae in the shallow part of the sea with sandy seabed can be differentiated from the algae that are found on rocky reefs.

The aim of the mapping of the underwater habitats is to develop a complete and detailed view of the underwater world through a multidisciplinary approach where hydrographers make a bathymetric map and analyze the acoustic characteristics of the bottom, geologists show lithology and sedimentology, while biologists observe underwater wildlife. In this way, the collected data is then combined into appropriately structured database that forms the basis for further interdisciplinary analysis. Based on the difference of perception in a certain period of time, it is possible to come up with many conclusions about the dynamics of the underwater environ-

ment. Accordingly, it allows the tracking of changes caused by human activities such as industry, pollution, increased traffic or certain methods of fishing, but also the impact of nature through waves, wind or ocean currents.

It is important to emphasize that the maps of the underwater habitats are in fact the best estimate of the spatial distribution of habitats, at some point, which was formed on the basis of currently available knowledge and skills at the disposal of experts. Therefore, this kind of mapping is still relatively limited by the process of improvement of the human knowledge on the underwater environment. Namely, there is no end point in the process of mapping the undersea habitats, since the maps created in such a way are a prediction, and need to be tested trough use. In a narrow sense, the mapping of the underwater habitats is the application of mapping the spatial distribution and extent of the habitat in the form of complete coverage of the sea floor (continuous surface) with indicative boundaries between adjacent habitats (Foster-Smith et al., 2007). Perhaps the most important consequence of the process of mapping underwater habitats is the fact that the maps that occurred in this way are not just a result of observation data, but the result of polemics about the distribution of habitats based on the best available information and opinions of experienced professionals.

This type of research is an extremely complex process and is different in each individual case. But nevertheless, we can identify the key and indispensable process steps:

- Optimal selection of location and sample density (sampling)
- Selection of the most appropriate method of observation (remote sensing)
- Integration and modeling
- Designing a map with respect to purpose (cartography)

When mapping the underwater habitat, data obtained by sampling and the data of continuous surface, containing physical characteristics of the habitat which are important for the clear separation between the classes of habitat, are combined. The latter can be determined directly from some form of remote sensing (e.g., acoustic techniques) or can be derived from the physical model. Ideally, the information should be derived from the developed models that directly link the physical characteristics of the habitat with biological data (Van Lancker and Foster-Smith, 2007). Because of the imperfections of mentioned models, experts often have to rely on their experience. The process of drawing conclusions about the spatial distribution of habitats is most often called, for lack of a better term, “modeling”. The objective of the integration of data obtained by different methods, or sensors, is that the required information of a demanding quality is determined with the maximum efficiency at minimum cost. The basic steps in the process of habitat mapping are shown in Figure 1.

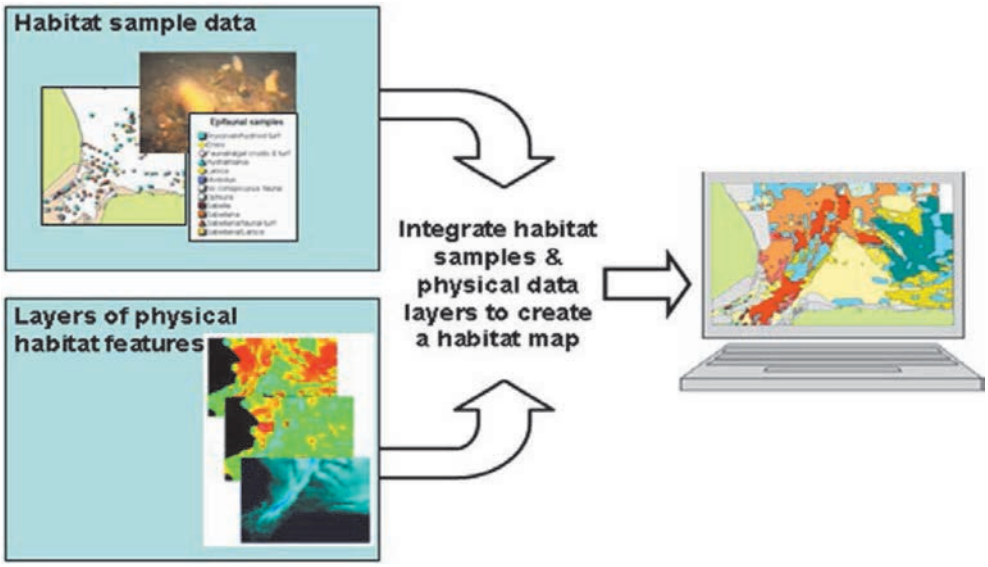


Fig. 1 – Basic steps in the process of habitat mapping (Foster-Smith et al., 2007)

Different methods of data collection for the purpose of mapping underwater habitats differ in quality, efficiency and economy. The term "data" generally refers to the numbers, however, one can use different types of written information, photos, videos, and physical samples (White and Fitzpatrick, 2007). The previous (classic) mapping methods were mainly based only on sampling and interpolation, whose resolution and efficiency cannot meet the needs of modern scientific management of underwater projects that require a full understanding of the dynamics of the underwater ecosystem. Furthermore, it should be noted that the observations are conducted under water, which makes the process of surveying more complicated than for terrestrial surveying, so the whole process is gaining importance as it is necessary to show the objects that are invisible to the human eye. In such conditions, the choice of remote sensors usually falls on the acoustic sensing technique because of the exceptional properties of the transmission of acoustic waves with water, as opposed to other properties (electromagnetic).

2.1 The principles of the underwater acoustics

Acoustics can be defined as the creation, transfer and reception of energy in the form of vibrating waves in matter (Pribičević, 2005). As such, today it is widely applied in systems whose task is underwater observation, since the water is compressible medium and allows the propagation of sound through it. Common acoustic phenomena are longitudinal acoustic waves. Passing along the media, the media particles vibrate and cause fluctuations in the density and pressure (acoustic pres-

sure) along the path of motion of the wave, thus, there is a movement of particles and changes in their distance due to the acoustic pressure.

The characteristic acoustic resistance is defined as the ratio of the acoustic pressure and particle velocity. The measure for the force that is transmitted through the acoustic wave is intensity I (W/m^2) (Pribičević, 2005). Accordingly, the intensity is the amount of energy that in one second flows through surface $1m^2$ in size, and is perpendicular to the direction of sound propagation. Acoustic intensities are usually in the range of 10^{-12} to 10^{1} W/m^2, thus a logarithmic scale of numbers is introduced to make it easier to express this range. Then we talk about the scale that represents the sound intensity (*IL*). The most commonly used is the scale of decibels (*dB*).

The largest number of systematic errors in the measurement of acoustic sonars applies to the determination of the exact speed of sound in different layers of water, since the velocity of propagation of acoustic waves depends on the characteristics of the media in which the wave spreads. In water, it depends on the two physical parameters (density and compressibility) that are functions of temperature, pressure (depth), and salinity, and it ranges from 1460 to 1560 *m/s*. Therefore, the speed of sound in water can be expressed as a function of three variables. There are several empirical formulas for determining the speed of sound in water from which it is clear that the speed of sound rises with the increase of any of the three variables. However, it should be noted that the greatest impact on the speed of sound in water has the temperature, so it is necessary to pay particular attention to the measurement of the speed of sound in the water column during the summer months in closed areas because of the pronounced warming of the surface layer of seawater.

The main factors that influence the spread of acoustic waves in water are: refraction, attenuation and reflection. Horizontal layering of the speed of sound in water, whereby the temperature decreases, and the pressure increases with depth, hinders the propagation of the acoustic signal so as to limit the observation of the seabed under the high vertical incident angles due to the occurrence of refraction. The refraction comes on the border between layers of different speed of sound and the acoustic wave is refracted toward a layer with higher-refractive index, that is the area with lower speed of sound. The second factor that affects the spread of acoustic waves is acoustic attenuation which can be divided into two key processes: absorption and scattering. Both processes lead to weakening of the acoustic signal. The superiority of acoustic signals under water in relation to electromagnetic is reflected in the depth of penetration of the signal into the water, but that does not mean that the absorption of acoustic signals in water is negligible. It is associated with physical and chemical properties of water and is one of the key factors in the perception of the seabed for the purposes of mapping of the underwater habitats. The second process of attenuation, scattering, occurs with multiple reflections of

acoustic signals due to different inhomogeneity that encounter on the path of the acoustic signal. The third factor, reflection, occurs upon arrival of the acoustic wave on the boundary surface between two different media or sea layers. The characteristics of the reflected signal depends on the incidence angle and the properties of the media that are provided as the characteristic acoustic impedance, which is the product of medium density and sound velocity in the same (Hansen, 2013):

$$Z_0 = \rho \times c \tag{2.1}$$

where c is the speed of sound in the medium, and ρ the density of the same.

Based on this, with underwater acoustic signals, just the opposite of complicated effects of electromagnetic waves, is easy to determine the differences between the various objects on the seabed and the water that surrounds them, as shown in Figure 2. Even slight changes in acoustic impedance within the sediment can be demonstrated in the measured data, which ensures successful implementation of the classification process.

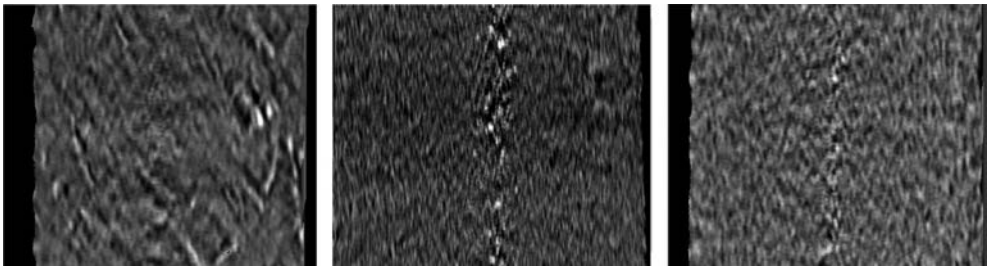


Fig. 2 – The samples of pseudo side-scan images (rocks, Posidonia beds, sand)

2.2 Basic principles of the backscatter theory

The sea bottom, unlike water, is highly resistant to the spread of the acoustic waves and does not allow relatively unobstructed propagation, as shown in Figure 3. The reason for this is the high density of the particles of the seabed, which disables free movements of the same. By the laws of physics, the total energy should be conserved. Obstructed part of the energy is reflected back into the water. Observed area reacts as a secondary source of acoustic waves. Scattered acoustic energy that is returned in the direction of the sonar is called the backscatter. Modern acoustic systems are able to register that energy, respectively the intensity of the backscatter, on which is possible to define features of the seabed. The echo level (*EL*) depends on the source level (*SL*), transmission loss (*TL*) and target strength (*TS*). This relation is given by the sonar equation (Lurton, 2010):

$$EL = SL - 2TL + TS\,, \tag{2.2}$$

Transmission loss is included twice for a double signal path, of the transducer to the seabed and back. This effect is caused by a circular spread of the acoustic signal.

Target strength (in dB) is defined as the ratio of the backscattered intensity (I_{bs}), and the incident intensity (I_i):

$$TS = 10log\frac{I_{bs}}{I_i}. \tag{2.3}$$

However, the target strength is the logarithmic expression for the backscattering and can be decomposed into two parts: the backscattering strength for a unit of surface (*BS)* in *dB*/m^2 and the actual ensonified area at the seafloor (backscattering area, *BA*) (Lurton, 2010):

$$TS = BS + 10log\left(BA\right). \tag{2.4}$$

The backscatter strength for a unit of surface represents the bottom reflectivity. It is often referred to as backscatter coefficient and can be expressed as (Jackson et al., 1986):

$$BS = \frac{R^2 I_s}{I_i A}, \tag{2.5}$$

Where A is the actual ensonified area, R is the slant-range, I_S is the scattered intensity and I_i the incident intensity. *BS* is dimensionless and therefore independent of the employed unit system. The backscatter strength is usually defined as the *dB*-value of the backscatter strength for a unit of surface and does not reference to a unit of length (Jackson et al., 1986):

$$\text{Backscatter strength} = 10logBS\,, \tag{2.6}$$

The backscatter strength varies in dependence of the seafloor characteristics and the incidence angle. The returned energy is inversely proportional to the incidence angle. So the backscatter strength will be high at small, and low at large incident angles (Lurton, 2010):

$$BS = BS_0 + 10logcos\theta, \tag{2.7}$$

where BS_0 represent the mean backscatter coefficient (mean unit backscatter strength) whose upper limit can be defined as around -5 dB/m^2 (Lurton, 2010). In practice, observed values for BS_0 are usually in the range of -10 to -40 *dB/m*^{2}, of course, depending on the terrain characteristics and sonar settings.

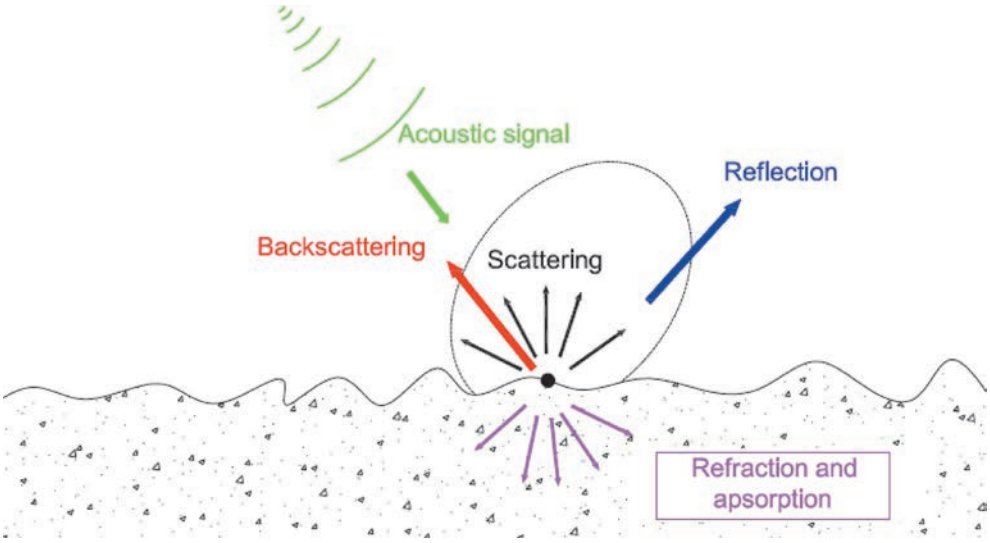


Fig. 3 – Acoustic signal backscattering

The intensity of the backscatter depends on the characteristics of the transmitted acoustic signal (incident angle, frequency), the distance from the target, the physical nature of the seabed (material, geometry, roughness) and its inner structure, as shown in Figure 4. Based on this it is possible to distinguish four types of backscatter. Interlaced backscatter is scattering due to differences between the characteristic acoustic impedance of the seabed and the water that surrounds it, and is a major component of the total backscatter. Volume backscatter depends on the heterogeneity of the seabed structure. Part of the energy that penetrates the sea floor (depending on the frequency and impedance contrast) is hampered by the heterogeneity of the sediments causing an increase in backscatter. As the incident angle is increasing, heterogeneity of the seabed has the greater impact in the total backscatter. The impedance of the seabed, among other things, depends on the grain size of the sediment. Therefore, the impedance is correlated to the roughness of the bottom sediment whose impact is defined as a Kirchhoff's backscatter. The larger the grains of sediment are, and surface roughness is higher, the impedance is increasing. Here we refer to the irregularities of the seabed whose dimensions are comparable to the wavelength of the acoustic waves. Finally, the intensity of the backscatter increases with the impedance contrast between water and sediment. By

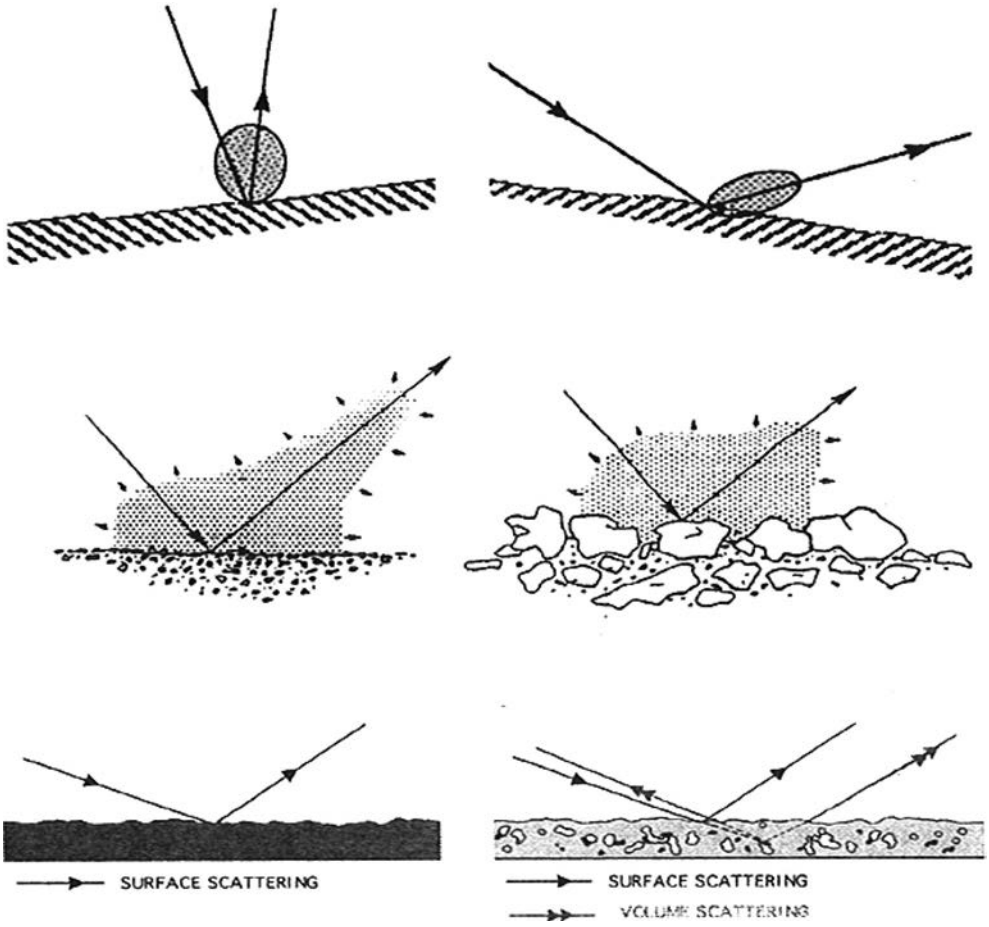


Fig. 4 – Types of backscatter due to a different sea bottom characteristics (Blondel and Murton, 1997)

mapping the underwater habitats, commonly observed value of the backscatter is total backscatter which is the sum of all the above mentioned types of acoustic backscatter signal.

In order to conduct classification of the seabed on the basis of the backscatter intensity, measurements should be first corrected for many different geometric and radiometric factors, and then can be compared with the model. This process is also called the Angular Range Analysis (*ARA*). Selected model should directly link the features of the seabed with the measured intensity (such as models by Jackson). Three main parameters which are modeling the curve on the basis of which classification is done are: acoustic impedance contrast, roughness and heterogeneity of the seabed.

2.3 Multibeam echosounders

Recent developments of the acoustic measurement techniques, especially of the multibeam echosounders (*MBES*), changed the way in which we can present and understand the ecosystems of the seabed. By using the MBES it becomes economical to observe really large areas of the seabed. Such studies provide basic data from which, integrated with data collected at the bottom of the sea, thematic maps of the seabed environment can be performed and interpreted. Multibeam echosounders are an expansion of singlebeam echosounders (*SBES*) which are transmitting only one vertical beam and the depth is determined based on the time measurements of double signal path. MBES simultaneously transmit several hundred beams down to the seabed. They can be divided into two groups: sweep and swath systems. Sweep systems consist of a row of closely placed SBES, mounted on the hull perpendicular to the vessel.

The Form of the system is quite complex because of the performance of the vessel's hull, and therefore the use of this system is largely limited to the port and the narrow channel (Pribičević, 2005). Swath system used in the preparation of this study produces multiple acoustic beams from one system transmitter. These systems allow rapid measurement means for determining the morphology and the nature of the sediments on the sea, river or lake bottom. Multi-element transducer allows many individual measurements of the water depth and power control for each ping. By the extremely large number of side reflected signals, their direction and distance record, and on this basis calculated depth, observation of extremely wide band in just one turn is allowed. Bundles of such systems often reach 8 times the width of the water depth in the area is observed. Thus, the greatest progress in relation to all other acoustic systems is in the efficiency of the system. The MBES measurements ensure sufficient accuracy required integration with additional sensors into a single measurement system. Therefore, the system should be capable of real-time import (import) and process data from other sensors (heading sensor, GNSS, SVP, motion sensor) during data collection. Taking into account all these factors, the vertical accuracy of the MBES in practice is better than 1% of the water depth (Lurton, 2010).

2.4 Pilot project of the seabed habitat mapping by analyzing the MBES backscatter

Practical application of the habitat mapping process, according to the popular methodology of conducting hydrographic survey, is divided into data collection and processing of the intensity of the return of the acoustic scattering and bathymetry. Before setting the test area we contacted Dr.sc. Ante Žuljević (Institute of

Fig. 5 – The DOF image of the Stari Grad bay with indication of the "test area" and "calibration area"

Oceanography and Fisheries, Laboratory for benthos, Split) in order to advise us with his expert advices and opinions. The test area is located on the southern coast of the Stari Grad bay on the island of Hvar, in the vicinity of the port (Figure 5).

This area was chosen because it was the first site of the invasive algae Caulerpa Taxifolia in Croatia, whose unstoppable expansion creates a real danger to the undersea ecosystem around the Adriatic Sea. Eradication was done manually with the aid of a suction pump and by covering the colonies with a black PVC foil with the purpose to prevent further expansion throughout the bay (Žuljević and Antolić, 2001). However, as a result of such proceedings, the seagrass meadows of Posidonia were destroyed (Figure 6). Because of its importance in the underwater ecosystem it is often called "the lungs of the sea". By the sampling frame, it was determined that in the pilot project area, Caulerpa taxifolia's habitat is still present. However, to a much lesser extent than it was before, and usually as a mixed habitat. On the relatively small sample of data the goal was to determine the possibility of mapping the areas that are covered by Posidonia in order to determine the stage of its restoration as a consequence of the process of destroying the other target species. The area in which it was possible to clearly allocate at least three different kinds of habitat is selected as test area.

In order to reliably classify the seabed from the data of the MBES it is necessary to have two groups of high-quality data. The first group is the sampling data collected in two ways. The process of sampling by diving was carried out two weeks before the observations of the study area. It was carried out by one of coauthors of this paper, Nino Pijanović, M.Sc., in cooperation with the "Diving center VIKING" from Hvar, after Dr.sc. Ante Žuljević gave the practical advices and recommended

Fig. 6 – Samples of Posidonia taken on the site

the sampling positions. Insight into the general state of the spatial distribution of underwater habitats within the area was carried out with the help of diving equipment to determine the schedule of sampling points. Sometime later, after finding the state of the monitored area, on randomly selected points, the physical samples of the seabed were collected. However, a more dense arrangement of sampling points was defined by the areas in the immediate vicinity of the indicative boundaries between classes, and at the very specific points in the context of underwater habitats (Figure 5). The above procedure is necessary for the classification of habitats in accordance with the actual spatial distribution of habitats, since the sampling data are compared directly with the data on the intensity of the backscatter, and then for evaluating the end product. Other types of samples are recordings of pseudo side-scan sonar system, built-in MBES system, collected during the survey (Figure 2). Displayed recordings were used for sampling between the points where the physical samples are collected with the purpose that some key features of the seabed would not remain unnoticed. These data represent an additional source of information, but also a kind of system of sampling quality control.

After proper calibration of the equipment the observation was conducted on the 13. July 2015. The used instruments were mounted on an aluminum speedboat Colnago 20s, as shown in Figure 7. The equipement consisted of: MBES Teledyne Reson SeaBat T20-P, positioning system Applanix POS MV WaveMaster and the sound velocity profiler Valeport miniSVP.

###

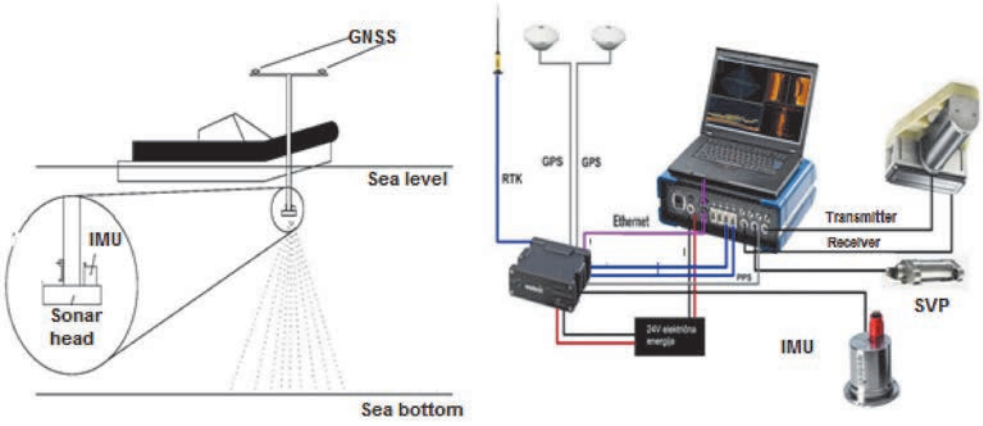


Fig. 7 – The principle of equipment installation (Tripodij d.o.o., 2015)

3. Results and Discussion

The research results can be divided into three main segments. The first is made of 3D terrain model based on the seabed, which provides the basis of geometric correction data on the intensity of the acoustic backscatter signal. The second segment represents adjusted backscatter data forming a mosaic in the field of research. Once created, backscatter mosaic should be compensated by the set of radiometric and geometric corrections (Time-varying gain correction, slant-range correction, Angular-varying gain correction, etc.) (Figure 8).

The analysis of statistical data of the associated values for each point of the mosaic, which are appropriately organized in the histogram, as shown in Figure 9, is

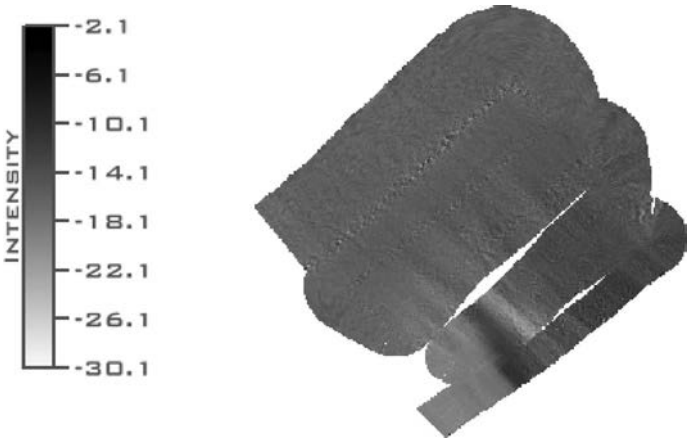


Fig. 8 – Corrected backscatter mosaic

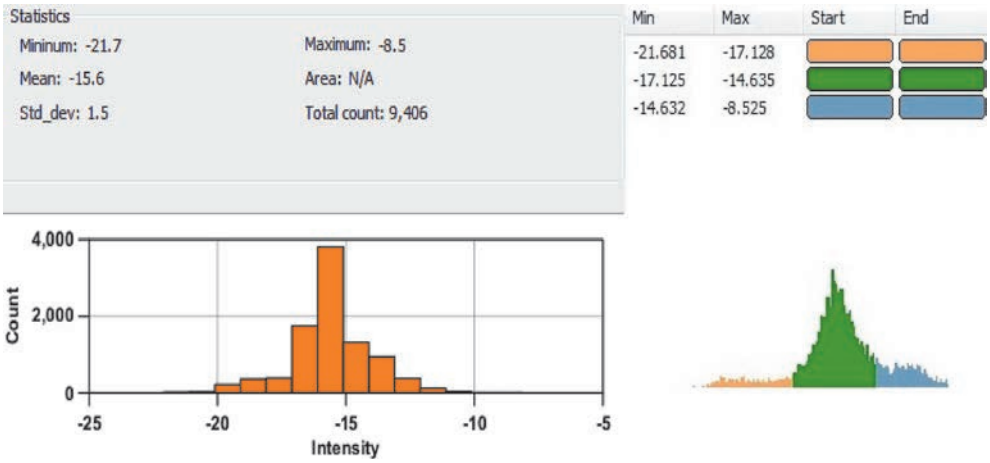


Fig. 9 – Statistical data of the classification

determined by the range of values for each class. By the comparison of the data sampling and the previously mentioned statistics, it was decided that the measured data indicates three classes.

By the previous action, all conditions for the development of the third segment, thematic map showing the spatial distribution of underwater habitats, were fulfilled. The first class represents the area covered in sand (orange), and is represented by the areas where the intensity of the backscatter was the lowest. Another class is the meadows of the seagrass Posidonia (green), whose backscatter intensity was

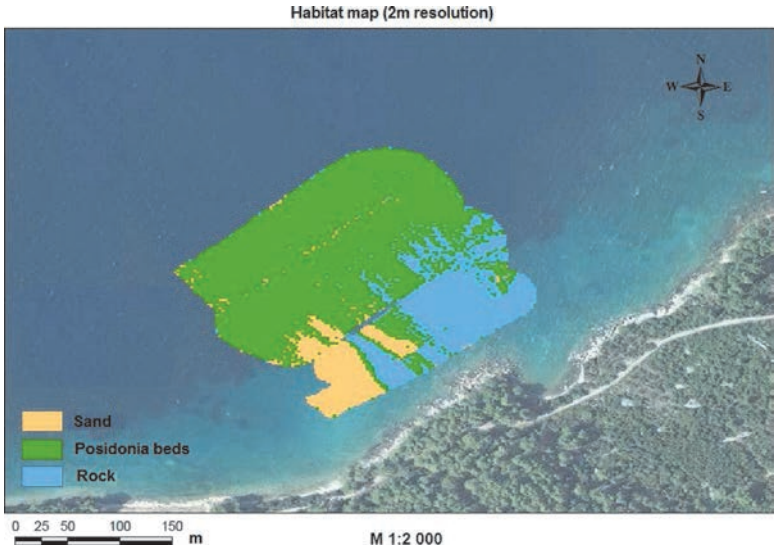


Fig. 10 – Habitat map crated for the test area

of medium size. The last class represents areas covered with rocks (blue), where represented areas had the strongest intensity of the acoustic backscatter signal.

The classification is made on the basis of the normal distribution and the resulting mosaic in the Figure 10 has a spatial resolution of 2 meters. Also, a 30 cm resolution mosaic, which provides additional information on the density of each class, was made. Processing of the data was carried out in the comprehensive bathymetric, seafloor imagery and water column data processing software *CARIS HIPS and SIPS 9.0*. It enables simultaneous processing of the multibeam, single beam, sidescan or LiDAR data. The latest technologies on 3D visualization are incorporated very efficiently in the software, processing making, and analysis for the hydrographic, oceanographic or any other marine environment. Implemented software tool *Geocoder,* which was created by Dr. Luciano Fonseca, is used to process the seafloor imagery data and analyze the intensity of the scattering of the acoustic signal (Fonseca and Mayer 2007), (Rzhanov et al. 2011).

4. Conclusions

Because of the MBES efficiency, the selection of the applied observation method proved to be the best possible option. Observations were carried out in ideal weather conditions, using highly modern and sophisticated instruments. In order to clarify the distribution of underwater habitats the 2m resolution mosaic, which proved to be an acceptable solution with regard to the spatial distribution of the above mentioned classes, was made. The standard deviation of the backscatter intensity is 1.5 dB, what gives the impression of a very reliable data. The lower resolution was the way to reduce the problem of stronger backscatter intensity near nadir. However, the problem of classification occurs at the edges of the area that is covered with rocks. Due to the impact of the volume backscatter and the heterogeneity of the sediment, the edges of rocks covered with sand belonged to the class of seagrass Posidonia which is not in accordance with the real distribution. The sand which was applied to the rock absorbed some of the energy and reduced the intensity of the rock's backscatter what resulted in placing these areas in the medium class which in this case represents Posidonia. Other boundaries between the classes are showing the spatial distribution of underwater habitats quite realistically and in accordance with the sampling. Given that the nature of the seabed is very dynamic, the two meter resolution distribution of habitats should meet the needs of most application. From the produced maps of underwater habitats in this pilot project, we can say with certainty that the damaged habitat of Posidonia, a few years after treatment in order to destroy the invasive algae, are almost fully restored, and that the habitats of Posidonia are very densely populated.

Based on the results it can be concluded that acoustic methods with reliable measurement of depth, certainly provide new opportunities to obtain quality information on the features of the seabed.

Current methods of habitat mapping exclude this type of remote sensing which is presented in the paper. Current methods of habitat mapping also have much lower resolution than this new methodology and need longer period of field work that solely depends on the experience of professionals who plan the sampling. Taking this into account, one can say that a significant progress is achieved in this field of research.

Acknowledgements

Special thank you to the companies “TRIPODIJ d.o.o” and “Geomar d.o.o” for the loan of equipment and instruments and for the advice and assistance during the observation. A big thank you to the company “CARIS HIPS and SIPS” on licenses for a software package that was used. Much obliged.

References

[1] Blondel, P. and Murton, B.J. (1997): Handbook of seafloor sonar imagery. 1st Edition, John Wiley and Sons, Praxis Publishing, Chichester.

[2] Fonseca, L. and Mayer, L. (2007): Remote estimation of surficial seafloor properties through the application angular range analysis to multibeam sonar data. Marine Geophysical Researches, p. 119-126.

[3] Foster-Smith, B., Connor, D. Davies, J. (2007): What is habitat mapping? In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough.

[4] Hansen, R. E., (2013): Introduction to sonar, Course material to INFGEO4310, University of Oslo, Oslo.

[5] Jackson, D., Winebrenner, D., Ishimaru, I. (1986): Application of the composite roughness model to high-frequency bottom backscattering. Journal of the Acoustic Society of America, 79(5), 1410-1422.

[6] Lurton, X. (2010): An introduction to underwater acoustics - principles and application. 2nd Edition, Springer, Berlin Heidelberg.

[7] Pribičević, B., (2005): Pomorska geodezija, Sveučilišni udžbenik, Geodetski fakultet Sveučilišta u Zagrebu, Zagreb.

[8] Van Lancker, V. and Foster-Smith, R. (2007): How do I make a map? In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough.

[9] White, J. and Fitzpatrick, F. (2007): How do I collect my data? In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough.

[10] Rzhanov, Y., Fonseca, L., and Mayer, L. A. (2011) : Construction of seafloor thematic maps from multibeam acoustic backscatter angular response data”, Computers and Geosciences Journal. Elsevier, Cambridge, MA, USA.

[11] Žuljević, A. and Antolić, B. (2001): Partial eradication of the Caulerpa taxifolia (Vahl) C. Agardh in Stari Grad Bay (Croatia) // Fourth International Workshop on Caulerpa taxifolia/ Gravez, V., Ruitton, S., Boudouresque, C.-F (ur.), Marseille: GIS Posidonie, 2001., p. 259-265.

Scientific paper

Effect of Embedded TiO_2 Particles into Electroplated Nickel Coatings on the Surface Roughness

Miroslav Gojo[1,*], Velizar Stanković[2,*], Vesna Grekulović[2], Tomislav Cigula[3]

[1]Croatian Academy of Engineering
[2]University of Belgrade, Technical Faculty Bor, Serbia,
[3]University of Zagreb, Faculty of Graphic Arts, Croatia

Nickel coatings onto brass substrate were produced from suspensions consisted of a conventional nickel sulphate bath and fine TiO_2 particles ($d_p < 1\mu m$) kept in suspension by magnetic stirrer. Composite coatings were obtained at different current densities and different concentrations of particles in suspensions while process time and temperature were constant.

Surface of composite coatings was characterized by means of optical and SEM micrographs, by EDS analysis and by determining surface roughness parameters. Results of the investigation showed that presence of the TiO_2 particles in the electrolyte causes an embedment of particles into deposited metal, changing physical characteristics of produced coating. Particles were embedded as smaller or bigger agglomerates rather than individual ones. Both varied parameters of electrochemical deposition, the current density and particles concentration in electrolyte, determine the amount of embedded particles into the coating thus affecting the surface quality. Certain irregularities appear on the surface in the form of whiskers and nodules at higher particles concentration as well as at higher current density. Furthermore, under these conditions, a certain rugosity of the surface was detected as a reflection to the electrolyte streamlines that exist due to the suspension stirring.

Surface roughness analysis has revealed trends in the measured values of amplitude roughness parameters. In the group of samples with composite coatings obtained at different current densities the highest values of measured roughness parameters were detected at 6 A dm^{-2} while the lowest values were obtained at 10 A dm^{-2}. As for group of samples with coatings obtained at different concentra-

*mirogojo@gmail.com, vstankovic@tf.bor.ac.rs

tions of TiO_2 particles higher values of measured roughness parameters were observed at lower values of TiO_2 concentrations. However, relationship between the profiles geometry and current density or concentration of particles in the bath has not been found.

Key words:

nickel; composite coatings; TiO_2 particles; surface roughness, optical and SEM analysis;

1. Introduction

Composite electrolytic coatings comprise a metal matrix into which fine inert, semi-conductive or conductive particles are incorporated from a plating bath containing corresponding metal ions and fine particles suspended in. Particles incorporated into metal matrix change the coating properties depending on the particles nature and on their concentration embedded into the coating. For example, hard particles will contribute to a coating hardness and wear resistance, while soft particles will result with a less hard – "softer" coating. Inert particles will improve corrosion resistance, while particles having catalytic properties will give a catalytic effect to the composite coating. Independently of the particles nature composite coatings have as a rule an increased surface roughness, what has not been studied to the extent it deserves.

In the recent time an increased attention has been paid to the electrochemically produced composite coatings, from different electroplating baths, with different nanoparticles incorporated in a various metal matrixes in order to achieve a beneficial either mechanical, electrical, magnetic, catalytic, anti-corrosive or optical features (Stanković, Gojo, 1996, Medeliene et al., 2003, Xue, et al., 2004, Tao et al., 2006, Roldan et al., 2007). Nickel composite coatings, electroless or electroplated, in combination with different kind of nanoparticles, have particularly attracted attention of scientists as nickel has been used for long time in plating of various metallic products used widely in household and construction. Beside protective and aesthetic properties of such coatings, nickel coatings are resistant against corrosion, but also have certain catalytic properties, particularly in hydrogen production technology, as it is Raney nickel. Because of these facts, nickel composite coatings have been most frequently studied in the past. Different combinations of fine particles and nickel containing baths served as a model-system in these studies, as: Ni-Al_2O_3; Ni-SiC; Ni-SiO_2,Ni-C (graphite); Ni-MoO_2; Ni-MoS_2; Ni-organic compounds either polymer particles or fine droplets of lubricants. Recently, an increased attention has been devoted in producing and further characterization of nickel composite coatings with TiO_2. Investigations had fundamental but applicative goals as well,

particularly for the use in photo-electrochemistry as well as to be used at higher temperatures (Thiemig, Bund, 2008, Spanou et al., 2009, Moustafa et al., 2013). Simultaneous embedment of TiO_2 and RuO_2 as well as Fe_2O_3 and TiO_2 into nickel matrix and the catalytic effect of such composite coatings on water electrolysis and hydrogen production has also been a subject of the recent investigations (Shilbi, Dilinon, 2007, Sebeelamol, 2013).

The aim of this study was to investigate the electroplating of Ni-TiO_2 composite coatings and the main parameters affecting the process The goal was to determine how the particles concentration in the bath, operating current density and other experimental conditions influence the amount of embedded particles and consequently on the coating surface quality.

2. Experimental

2.1 Materials and Methods

Eletrodeposition of nickel composite coatings have been performed in a laboratory cell (volume 600 cm^3), equipped with magnetic stirrer and heater. The electrodeposition were was performed by using a commercial bath for nickeling (producer's mark Ni-EX), provided by Metal Salt Factory, RTB Bor, Serbia. The bath was ready to use, meaning that beside nickel sulphate solution of defined concentration of Ni^{2+} ions, it contains all additives for producing bright and ductile nickel electrocoatings. Solutions for surface degreasing, as well as for deoxidizing were both delivered by the same supplier.

TiO_2 particles, producer mark: RC8/1, No 145, was supplied by "Cinkarna" Celje, Slovenia. The average size of particles was in the range of 200 to 350 nm. Having a tendency towards agglomeration, a small fraction of bigger or smaller agglomerates are usually formed, with a size < 1 μm. This fact, associated with the particles wettability, led to the need of a longer time of stirring the suspension prior switching the cell on, in order to achieve as much as possible homogeneous suspension with a minimum agglomerates.

Electrodeposition of composite coatings was performed onto brass plates (1.5 x 1.7 cm), as cathodes, placed among two anodes made from pure, rolled nickel, forming a side-by-side configuration. As a DC current supplier "Amel" Potentiostat-galvanostat, 555B was used. Operating current and cell voltage were monitored by ammeter and voltmeter while pH by a pH-meter.

2.2 Experimental procedure

Prior to the electroplating, the brass cathode was degreased and deoxidized by using corresponding solutions for degreasing and deoxidizing in accordance with the instruction for use, given by the supplier. After each of these preparative operations, the electrode was rinsed thoroughly with distilled water. The electrolyte was prepared by adding a corresponding amount of TiO_2 particles in the nickel bath while stirring at 600 rpm, in order to form a suspension and to keep all added particles in suspension. Before the electroplating, suspension was warmed up in order to achieve an operating temperature of 55 °C. When suspension was ready, prepared cathode was immersed into the suspension among the anodes and fixed to keep the same and stable distance between them. Electrodes were connected with the DC supplier; the operating current was chosen and adjusted to correspond to a desired current density; the cell and chronometer were both switched on at the same time and the electroplating process starts. Electrodeposition time kept constant in all experiments and was 70 minutes. Suspension concentration and current density were the operating variables. When the electrodeposition was completed, the cell was switched off; the cathode was removed from the electrolyte, rinsed several times with distilled water and then with ethanol, dried and stored. The electroplated electrode was cut in smaller peaces to be used in further characterization.

The following methods and equipment were used in the produced composites characterization:

- Optical microscopy (Olympus BX51 metallurgical microscope)
- Scanning electron microscopy with Energy-dispersive X-ray spectroscopy EDS (SEM, JEOL JSM-6460)
- Surface roughness (PERTHOMETER S8P Mahr GmbH).

Measurement of surface roughness was performed by using stylus instrument Perthometer that enables two-dimensional tracing of a surface. Traceability of used measuring instrument is assured using certificated calibration artefacts, i.e. Croatian national roughness standards.

Determination of amplitude roughness parameters *Ra* and *Rz* used in this study is compliant to the geometric product specification standards (ISO 4287:1997, ISO 4288:1996, ISO 3274:1996). The used measuring unit is equipped with the diamond stylus of 5 μm radius. During the measurement the stylus was moved at a constant speed across the samples with a measuring force of 1.3 mN. Measurements were performed using Gaussian filter with cut-off $\lambda c = 0.25$ mm and the evaluation length $ln = 1.25$ mm. The stylus is traversed normal to the surface at constant speed.

3. Results and Discussion

3.1 Surface roughness analysis

Three profiles of roughness parameters were measured on each sample. Obtained results are given in Tables 1 and 2, where $\overline{Ra}$ and $\overline{Rz}$ are arithmetic means of the three repeated measurement, and s is the standard deviation of these results.

Table 1 – Measured roughness parameters for the samples obtained at different current densities

TiO_2 Concentration = 10 g dm^{-3}				
Current Density, A dm^{-2}	**$\overline{Ra}$, µm**	**s, µm**	**$\overline{Rz}$, µm**	**s, µm**
2	0.376	0.093	1.841	0.236
4	0.367	0.063	1.746	0.342
6	0.162	0.020	0.716	0.028
10	0.619	0.100	3.662	0.708
20	0.203	0.042	1.103	0.151

Table 2 – Measured roughness parameters of the samples obtained for different TiO_2 particles concentration

Current Density = 10 A dm^{-2}				
TiO_2 Concentration, g dm^{-3}	**$\overline{Ra}$, µm**	**s, µm**	**$\overline{Rz}$, µm**	**s, µm**
10	0.315	0.055	1.515	0.056
20	0.440	0.047	1.911	0.146
40	0.160	0.005	0.863	0.014
80	0.165	0.005	0.810	0.113

Recorded roughness profiles did not reveal any characteristic or common geometry within a group of samples with constant current densities i.e. concentrations of particles, as well as between those two groups of samples, (Fig. 1 and Fig. 2).

Surface roughness analysis has revealed trends in the measured values of amplitude roughness parameters. In the group of samples with composite coatings obtained at different current densities lowest values of measured roughness parameters were recorded at 6 A dm^{-2} and highest values at 10 A dm^{-2}.

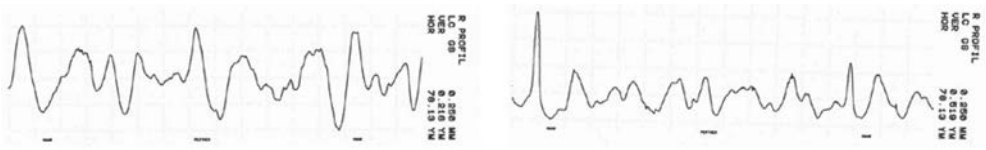


Fig. 1 – Profilograms across the composite coatings surface: TiO_2 Concentration = 10 g dm^{-3}

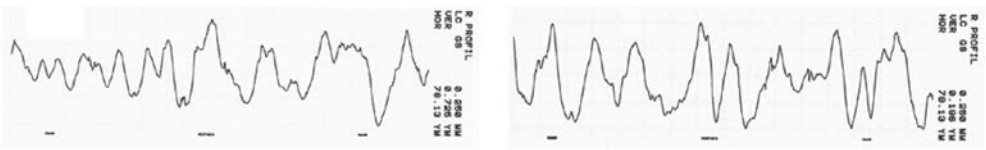


Fig. 2 – Profilograms across the composite coatings surface: Current Density = 10 A dm^{-2}

As for group of samples with coatings obtained at different concentrations of TiO_2 particles higher values of measured roughness parameters were observed at lower values of TiO_2 concentrations, table 2.

3.2 Optical, SEM and EDS analysis of the surface

3.2.1 Effect of current density

Optical (a) and SEM (b) images of the surface of nickel composite coatings are presented in Fig. 3, for the samples obtained at constant TiO_2 concentration but different current densities in the electroplating and in Fig. 5, for the samples produced at constant current density but at different TiO_2 concentrations. It could be seen in Fig. 3(a), that higher current densities corresponds to the smoother and more uniform coating surface with smaller or bigger nodules visible under optical microscope. Besides single TiO_2 particles built-in the coating, bigger and smaller agglomerates of TiO_2 are detectable onto the coating surface as shown on the SEM photos in Fig. 3(b). Moreover, higher current densities corresponds denser appearance of TiO_2 particles on the surface, as is visible the SEM micrograph, Fig 3(b) right. Also, higher current densities corresponds faster metal ions reduction rate, causing finer nickel grains of deposited metal, as well as stronger particles attraction to the cathode surface that occurs at higher cathode potentials. High metal deposition rate coupled with stronger attraction of TiO_2 particles leads to an increased amount of particles built-in the composite coatings. It also means that at higher current densities corresponds higher surface density of embedded particles either as single ones or as agglomerates.

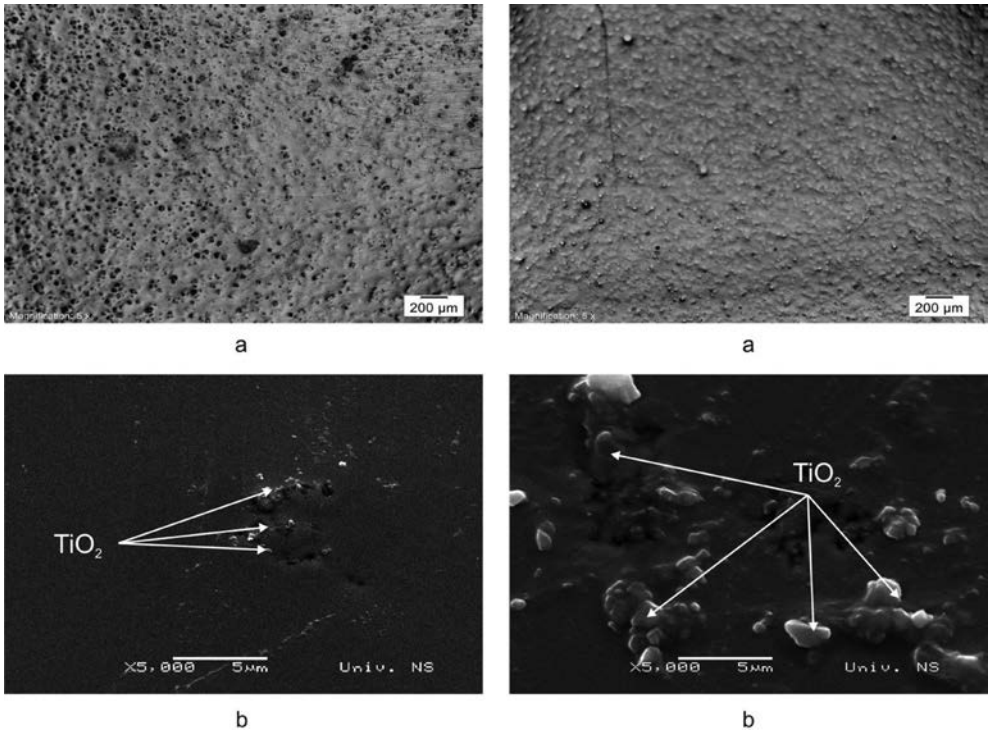


Fig. 3 – Optical (a) and SEM images (b) of nickel- TiO_2 composite coatings surface

These observations about the surface smoothness and TiO_2 particles uniformity across the surface are in accordance with the recorded profilograms and data evaluated from them (See Table 1 as well as Fig. 1 and Fig. 2). Considering the contribution of current density to the surface quality, it could be seen that more nodules were formed at lower as well as at higher current densities, while minimal percentage of TiO_2 particles on the surface was detected at 6 A dm^{-2}.

Inclusions that appear on the surface of composite coatings (Fig 3.b) were analyzed by means of EDS and corresponding spectra are presented in Fig. 4 (b), detecting only nickel and titanium as constituent elements of the coating.

The EDS analysis was made on the parts of the surface on which agglomerates are seen, but the surface of the composite coating without deformations, i.e. visible agglomerates as well. The analysis showed that agglomerates contain only nickel and titanium (photo Fig 4b left, spectrum 2), while the spectrum 3 taken for the flat surface without agglomerates or single particles shoes only nickel. Similar result could be seen in Fig 4b, right, where spectrum 1, taken for agglomerate consists from Ni and Ti, while spectrum 2 detected only Ni. The particles of TiO_2 are certainly built into the composite coating and the amount of the particles in the com-

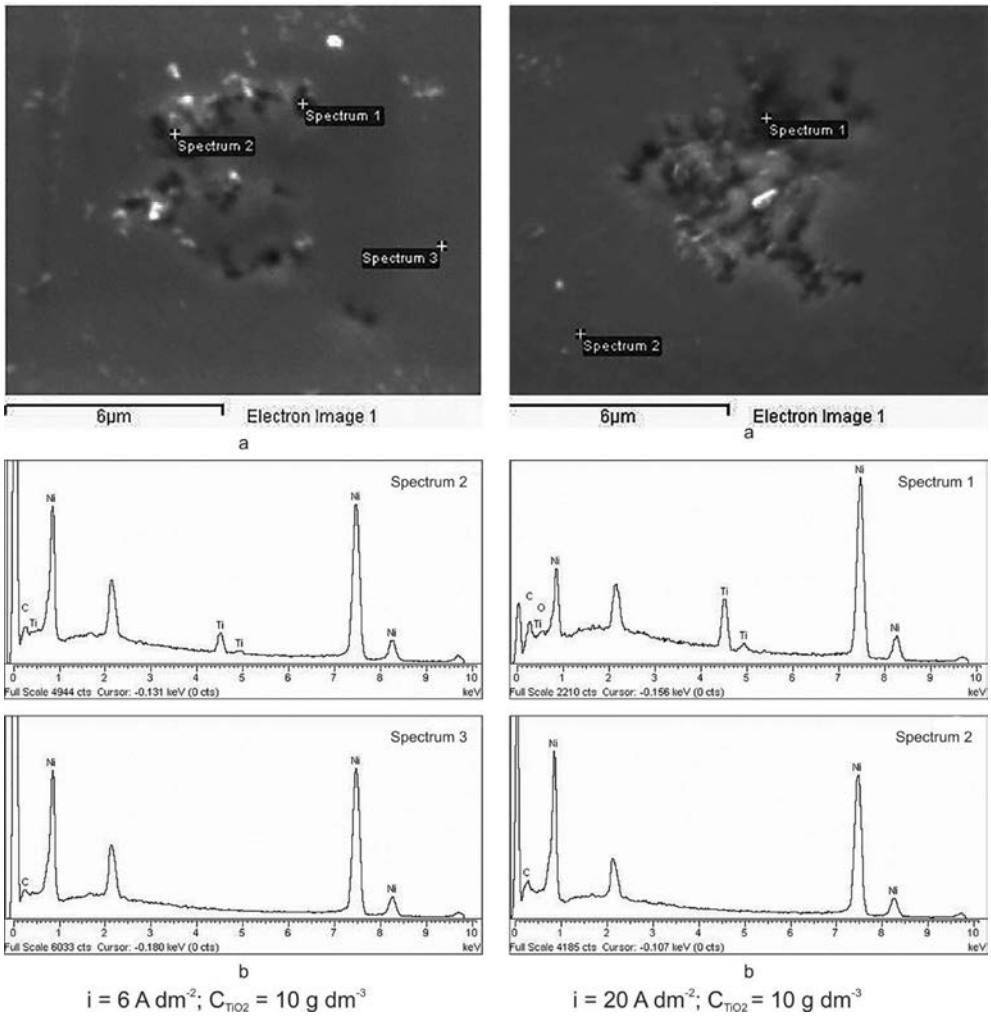


Fig. 4 – EDS analysis: SEM image of a detail of the surface (a), EDS spectra (b)

posite coating is dependent on the current density and the concentration of TiO_2 in the electrolyte, what was confirmed by gravimetric analysis (M. Gojo, et al., 2011).

Plotting the fraction of included TiO_2 particles against the current density, as presented in Fig. 5, the fraction of TiO_2 occluded by nickel in agglomerates of the composite coating surface could be evaluated. Obviously higher current densities (higher electrode potentials), corresponds to an increase in number of TiO_2 particles that would be attracted to the electrode to be occluded on the surface as a single one or in the form of agglomerates.

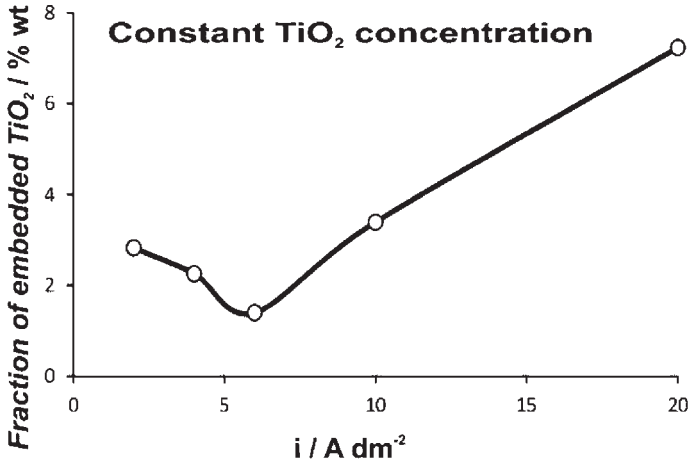


Fig. 5 – Fraction of embedded TiO_2 particles on the analysed micro-surface vs. current density: C_{TiO2} = 10 g dm^{-3}

Significant increase in the TiO_2 fraction on the surface occurs at current densities higher than 6 A dm^{-2} (Fig. 5). Again, it is confirmed that TiO_2 agglomerates that appears as nodules on the surface contain a significantly higher TiO_2 content than the rest of surface what affects a pronounced surface roughness.

3.2.2 Effect of particles concentration in the bath

Keeping the current density at 10 A dm^{-2} and changing the concentration of TiO_2 in the suspension, a series of coatings were obtained and the surface of which was analyzed in the same way as it has done previously. The results are presented in Fig. 6, as optical (a) and SEM images (b). Optical photos have shown that higher concentration of TiO_2 in the bath results with smoother surface. That also can be seen from Table 2.

SEM photos shown the existence of nodules grouped as islands onto the surface. By them, these nodules are smaller on the samples obtained from the bath containing 80 g dm^{-3} of TiO_2 particles. Flatter surface with smaller agglomerates could be attributed to an attrition effect of the suspension to the electrode. Higher concentration of particles in suspension has a polishing effect, leading to smoother and brighter composite coatings.

Similarly, like in the previous consideration, the EDS spectra have shown nickel and titanium in case of spectrum 1, relating to agglomerates, while only nickel for

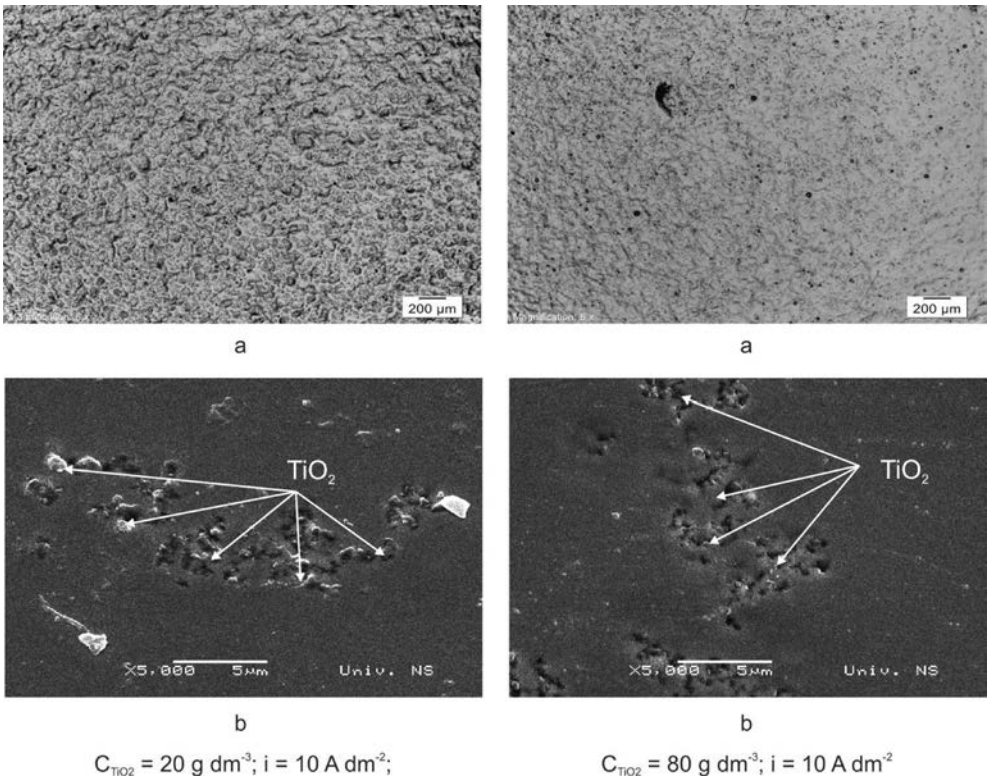


Fig. 6 – Optical (a) and SEM images (b) of Ni-TiO_2 composite coatings surface

spectrum 2, as has shown on the photos (Fig. 7), taken for two particles concentrations of the bath.

The fraction of TiO_2 in the coating, plotted against the particles concentration in the suspension, as could be seen in Fig. 8, has shown that there is no significant difference in the titanium percentage built into the composites produced from the baths containing the lowest and the highest concentration of particles. The existed minima of TiO_2 particles embedded in the composite coatings that appear in Figs. 5 and 8, need to be deeper investigated and explained. The difference is in the surface quality. It could mean that the composites, deposited from more concentrated suspensions contain approximately the same amount of included TiO_2 as those produced from eight times lower concentration. In the second case, embedded TiO_2 particles are in the form of bigger or smaller nodules rather than to be in the form of single particles homogeneously distributed through and across the coating. The question only is what kind of the surface quality we need – to be smoother or rougher.

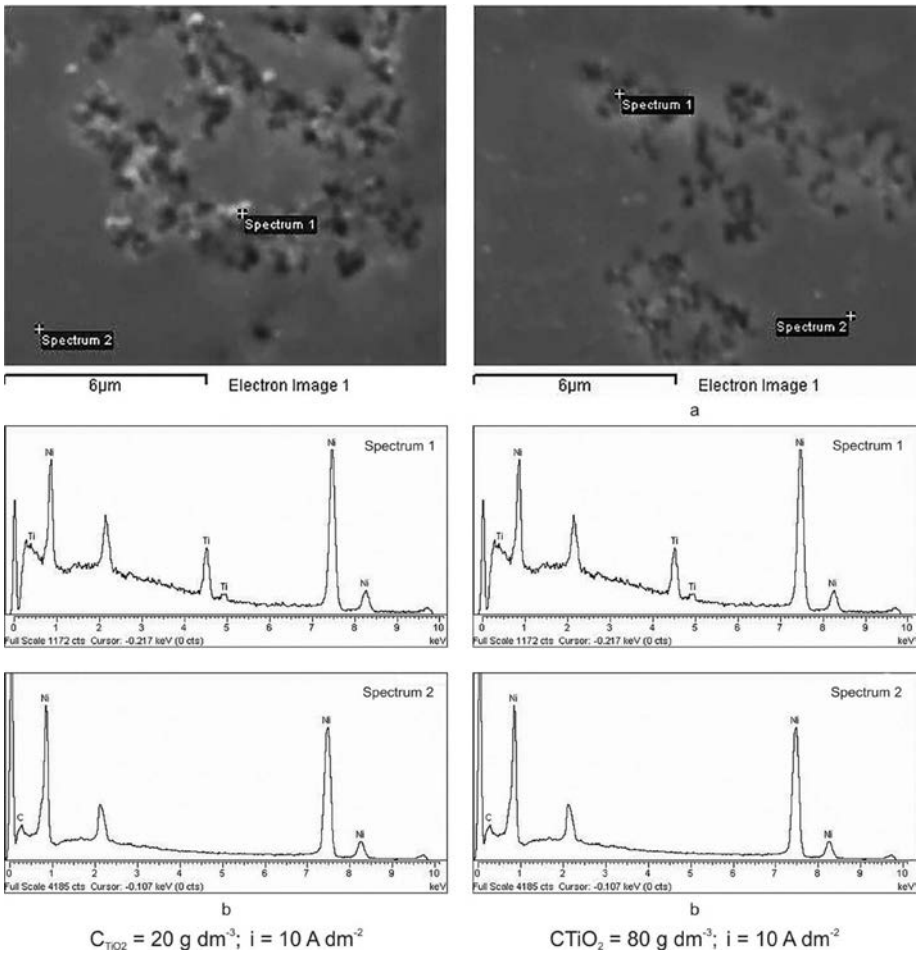


Fig. 7 – EDS analysis: SEM image (a), EDS spectra (b)

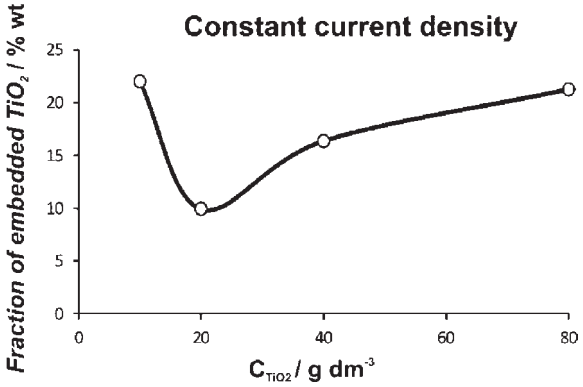


Fig. 8 – EDS analysis – Fraction of TiO_2 in coatings plotted against TiO_2 concentration in the bath: current density 10 A dm^{-2}

4. Conclusions

Composite coatings of nickel and TiO_2 were produced by electrodeposition of nickel from bath in which TiO_2 particles were suspended. During the electroplating, TiO_2 particles were attracted to the electrode surface where they were captured and incorporated into metal matrix. The amount of, in such a way embedded particles, depends on the applied current density and the particles concentration in the bath. In both cases the dependences are complex having well expressed minima on the graphs. Titania particles, included in the metal matric cause an increased surface roughness due to a formation of small nodules on the composite coatings consisting from smaller or bigger aggregation of TiO_2 formed onto the surface. Nodules are more expressed at lower current densities and, particularly at lower TiO_2 concentrations in the suspension.

Surface roughness analysis has revealed trends in the measured values of amplitude roughness parameters. In the group of samples with composite coatings obtained at different current densities the highest values of measured roughness parameters were recorded at 6 A dm^{-2} while the lowest values were obtained at 10 A dm^{-2}. As for group of samples with coatings obtained at different concentrations of TiO_2 particles higher values of measured roughness parameters were observed at lower values of TiO_2 concentrations.

At higher concentrations the surface is smoother, with smaller or even no nodules despite an increase in the amount of embedded particles were observed by SEM images on the coating surface. This effect is attributed to an attrition effect of particles in motion over the electrode surface, resulting in a polishing effect. EDS spectra show that only Ni and Ti are detected in the nodule under consideration by EDS analysis. In surface regions farer from nodules only Ni was detected.

Acknowledgements

The authors want to express their sincere thanks to the Metal Salt Factory, RTB Bor, Serbia for providing us the chemicals used in the experiments as well as to the titania producer Cinkarna Celje, Slovenia, for providing us the titania samples. The authors also thank to Miss Goranka Baršić, University of Zagreb, *Faculty of Mechanical Engineering and Naval Architecture, Croatia* at the roughness measurements, *and* Mr Miloš Bokorov, Faculty of Science, *Department of Biology and Ecology*, University of Novi Sad, Serbia, at the SEM and EDS analysis.

References

[1] V.D. Stanković, M. Gojo, Electrodeposited composite coatings of copper with inert, semiconductive and conductive particles. Surf. Coat. Technol., 81, 225-232 (1996), Elsevier

[2] V. Medeliene, V.Stankovic, G. Bikulcius, The influence of artificial diamond additions on the formation and properties of an electroplated copper metal matrix coating. Surf. Coat. Technol. 168, 161-168 (2003), Elsevier

[3] A. Roldan, E. Gomez, S. Pane, E. Valles, Electrodeposition of copper–magnetite magnetic composite films. J. Applied Electrochem.; 37, 575-582 (2007), Springer

[4] Y.J. Xue, D. Zhu and F. Zhao, Electrodeposition and mechanical properties of Ni-La_2O_3 nanocomposites. J. Mater. Sci. 39, 4063-4066 (2004), Springer

[5] S. Tao D. Y. Li, Tribological, mechanical and electrochemical properties of nanocrystalline copper deposits produced by pulse electrodeposition. Nanotechnology 17, 65-68 (2006), Hindawi Publishing Corporation

[6] D. Thiemig and A Bund; Characterization of electrodeposited Ni-TiO_2 nanocomposit coatings. Surface and Coating Technol. 202, 2976-2984 (2008), Elsevier

[7] E. M. Moustafa, A. Dietz, T. Hochsattel, Manufacturing of nickel/nanocontainer composite coatings. Surface and Coating Technol., 216, 93-99 (2013), Elsevier

[8] S. Spanou, E. A. Povlatou and N. Spirellis; Ni/nano-TiO_2 composite electrodeposits: Textural and structural modifications. Electrochim. Acta 54, 2547-2555 (2009), Elsevier

[9] S. M. A. Shilbi, V. S. Dilinon, Development of TiO_2-supported nano-RuO_2-incorporated catalytic nickel coating for hydrogen evolution reaction. Int. J. of Hydrogen Production; 33, 1104-1111 (2007), Inderscience publisher

[10] J. M. Sebeelamol, Development of Fe_2O_3–TiO_2 mixed oxide incorporated Ni–P coating for electrocatalytic hydrogen evolution reaction. Int. J. of Hydrogen Production; 38, 2271-2282 (2013), Inderscience publisher

[11] Gojo, M., Cigula,T., Pajkić, N., Stanković, V. (2011)., Euroanalysis, "Challenges in Modern Analytical Chemistry, Abstracts, USB AM1, June 2011, Beograd, Serbia

[12] ISO 4287:1997 Geometric Product Specification (GPS) – Surface texture: profile method–terms, definitions and surface texture parameters

[13] ISO 4288:1996 Geometric Product Specification (GPS) – Surface texture: profile method–rules and procedures for the assessment of surface texture

[14] ISO 3274:1996 Geometrical Product Specifications (GPS) – Surface texture: Profile method – Nominal characteristics of contact (stylus) instruments

Review paper

Progress in Polymerization of Olefins and Polyolefin Materials

Zvonimir Janović*, Ante Jukić*

Faculty of Chemical Engineering and Technology, University of Zagreb,
Marulićev trg 19, 10000 Zagreb, Croatia

This article reviews the remarkable progress in α-olefin polymerization, particularly of the catalytic systems with a simultaneous breakthrough in production processes and the improvements in the properties and application possibilities of polyolefins. Polyolefins are mostly ethylene and propylene homopolymers, copolymers, blends and composites. Due to their diverse and desirable properties and good balance of physical and chemical properties, their consumption has rapidly and significantly increase. It started with the discoveries in 1933 by E. W. Fawcett and R. O. Gibson of ethylene free radical polymerization under very high pressure and temperature. The real breakthrough was made by the discovery of organometallic heterogeneous catalysts, very efficient for α-olefin polymerizations. Their advantages are mild process conditions, high yields on obtained polymers and their desirable macromolecular structure. They resulted from the discoveries by Karl Ziegler and stereospecific polymerization by Giulio Natta. After them new catalysts were named *Ziegler-Natta (Z-N) catalysts*, and both inventors were awarded the Nobel Prize in 1963. The research on Z-N catalysts has been intensive and continuous all over the world, up to present time, leading to their improvement and great development of polymer industry. In the period 1980 – 1990 a new class of organometallic but homogeneous catalysts known as *metallocenes* based on complexes of bicyclodienyl and transition metals followed. Also, appropriate co-catalysts were discovered. The late development has led to new complexes, known as *post-metallocene* catalysts. Their major advantage is high solubility and activity. They enable the easy tailoring of polyolefins of different structure and morphology including copolymers with polar monomers. A huge breakthrough that followed in polymerization pro-

*Croatian Academy of Engineering member emeritus; zjanov@fkit.hr, ajukic@fkit.hr

cesses such as fluid bed, liquid phase loop reactor and reactor granule technology led to significant development and growth of polyolefin production. Also, the new materials, such as functional copolymers, bimodal polymers, and polyolefin nanocomposites have been made.

Key words:

polyolefins, Ziegler-Natta catalysts, metallocene catalysts, Spheripol process, polyolefin nanocomposites

1. Introduction

Polymeric materials belong to the most important classes of the presently used technical materials. Their production and utilization have made tremendous progress in the last fifty years and their use is ever increasing. Among them, polyolefins are the most applied and one of the fastest growing polymers regarding their good balance of physical and chemical properties, the high technology and sustainability of the polymerization processes, their excellent thermomechanical properties, good environmental capability including easy processing and recycling combined with their low cost and large scale production. They consist mostly of several α-olefin monomers, primarily of ethylene, propylene, their copolymers, blends and composites and in a less extent 1-butene, isobutene, 4-methyl-1-pentene and higher α-olefins. The variety of polyolefins produced by polymerization of these monomers include large volume of several types of polyethylene (PE) and polypropylene (PP) and specialty polymeric materials such as ethylene-propylene elastomers (EPR), ethylene-propylene-diene elastomers (EPDM), poly(1-butene) (PB-1) and some others. (Pritchard, 2012)

The history of polyolefins starts with the synthesis of long methylene chains, „polymethylene“ in 1879 by decomposition of diazomethane by German scientist, H. von Pechmann. In 1933 at Imperial Chemical Industry Co. (ICI) in England, E. W. Fawcett and R. O. Gibson found a polymerization process of ethylene into high molecular weight PE, which was performed under very high pressure (up to 2000 bar) and temperature (up to 250 °C) initiated by trace amounts of oxygen. It became the basis for industrial production of low density PE (LDPE). Its first commercial application came during the 2nd World War as ideal electrical insulation material for the new radar equipment. Afterwards in early 1950s, chromium-oxide based catalyst for ethylene polymerization was discovered by J. P. Hogan and R. L. Banks at Phillips Petroleum Co. in USA. The catalyst was highly effective for the ethylene polymerization at the lower pressure and temperature than in the ICI process, giving high density polyethylene (HDPE), but there was no possibility to produce polypropylene. (Hogan et al. 1986)

Karl Ziegler, a German scientist at Max Planck Institute in Mülheim, in 1953 discovered a new process for the polymerization of ethylene into linear polyethylene under mild conditions by using titanium chloride and alkyl aluminum catalytic system that was superior to all the existing ones. (Mülhaupt, 2003) Giulio Natta, an Italian scientist at University of Milan, in 1954 obtained for the first time isotactic polypropylene and Montecatini Co. started its production already in 1958. He conducted pioneering studies on the chain microstructure of synthetic organic polymers and postulated the mechanisms of stereospecific polymerizations. These historical discoveries are together called Ziegler-Natta (Z-N) catalysts and Z-N polymerizations. (Busico, 2013)

The discoveries of organometallic catalysts and stereospecific polymerization of α-olefins, dienes and a number of vinyl monomers by Karl Ziegler and Giulio Natta, their developments and recent progress is one of the most important achievements in the field of catalysis, macromolecular science and polymer materials in the last century. Their inventors were awarded the Nobel Prize in Chemistry in 1963 "for their discoveries in the field of chemistry and technology of high polymers" (Figure 1). These discoveries have stimulated an intensive, both basic and applied research all over the world, up to the present times, leading to great development of the polymer industry. (Gahleitner et al. 2013)

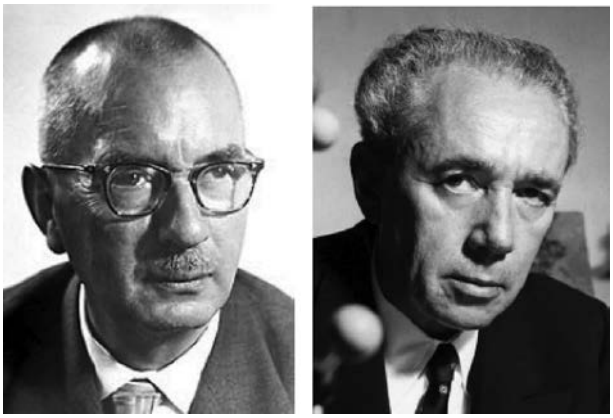


Fig. 1 – Nobel Prize Winners in Chemistry 1963: Karl Ziegler (left) and Giulio Natta (right)

In the 1980s, a new family of transition metal complexes known as metallocene catalysts was discovered, based on biscyclopentadienyl and transition metals such as Zr, Hf and Ti and appropriate cocatalysts. Their major advantage is solubility, high catalytic activity, and the easy tailoring of the polymer microstructure. (Schiers et al. 2000) The development of catalysts based on diimine complexes of nickel and palladium, and of phenoxyimino complexes of zirconium and nickel, resulted in polyolefins of different structure and morphology, and many new copolymers of

ethylene, particularly with polar monomers, producing a variety of new functional polymers, reactive oligomers, and block copolymers. Commercial production of polyolefins was utilized for the several main process types: high-pressure, solution phase, gas phase, and slurry phase. The breakthrough in polymerization processes such as fluid bed, liquid phase loop reactor and reactor granule technology led to significant development and growth of polyolefin production. (Severn et al. 2005) Despite the outstanding properties of polyolefins, the addition of some fillers and/or reinforcements such as silica, alumina, magnesia, calcium carbonate, titanium carbide, silicon carbide, montmorillonite, etc. are used to improve further some of their mechanical and physical-chemical properties such as gas permeability, biodegradability, flammability, etc. The preparation of polyolefin nanocomposites, particularly with carbon nanofibers, nanotubes and graphene seem to be the most important method to obtain novel properties and characteristics. (Kaminsky, 2012)

2. Polyolefin materials

Polyolefins can be divided into two main classes: polyethylene and polypropylene, which is than subdivided into several types for different applications. Their macroscopic properties strongly depend on the structure of macromolecular chain and composition in the case of the copolymers. (Whiteley et al. 2005) Polyethylene is the most popular and widely used polymer to date. It has very simple structure consisting of linear macromolecules of high molecular weight having high chain flexibility and tendency to crystallize. However, its macromolecules contain also some short chain branches (SCB) and in a less extent long-chain branches (LCB) as defects in polymer chain, which decrease the degree of crystalline fraction. Thus, the morphology of polyethylene is changed from chain-folded lamella to amorphous structure with increasing degree of short chains (Figure 2, adopted from Mülhaupt, 2003) and this decreases the density and melting temperature. In this way density is a reflection of degree of SCB in the PE chain. It, however, also depends to a degree on its molecular weight, all other factors being the same. Poly-

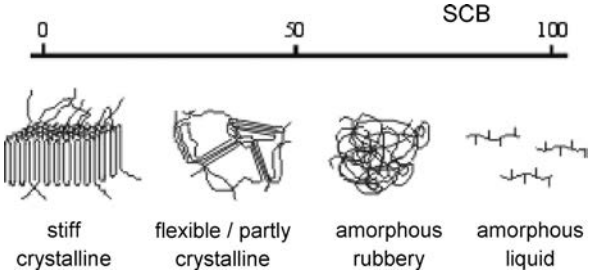


Fig. 2 –. Polyethylene morphology changes with a degree of short chain branching (SCB)

ethylene with high molecular weight tends to have a slightly lower density than that with low molecular weight average.

According to the density range, polyethylene polymers are classified into three main types: low-density PE (LDPE, densities from 0.915 – 0.940 g cm^{-3}), linear low-density PE (LLDPE, densities from 0.915 – 0.940 g cm^{-3}), high-density PE (HDPE, densities from 0.945 – 0.970 g cm^{-3}). Also, in minor amount very low density PE (VLDPE, 0.88 – 0.915 g cm^{-3}), medium density PE (MDPE) and ultra-high molecular weight PE (UHMWPE) are produced as a type of high density PE with molecular weight averages of several millions. Due to the broad applications of polyolefins the world production of PE and PP was 151 million tons in 2014, that is 63 % of all plastomers (240 million tons) and together with thermosets, elastomers and fibers (all together) about 300 million tons of organic synthetic polymers (Figure 3). (adopted from Shamiri et al. 2014)

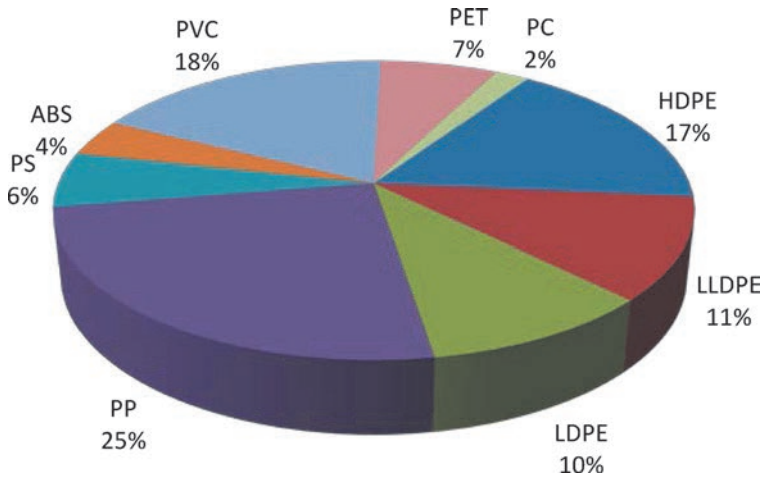


Fig. 3 –. The share of thermoplastic materials in world (2014) for total production of 251 million tones: HDPE – high density polyethylene, LLDPE – linear low density polyethylene, LDPE – low density polyethylene, PS – polystyrene, ABS – acrylonitrile/butadiene/styrene terpolymer, PVC – polyvinyl chloride, PET – polyethylene terephthalate, PC – polycarbonates.

3. Olefin catalytic polymerization

Apart from LDPE, which is prepared by radical polymerization at high pressure and temperature, all of these polyolefins (HDPE, LLDPE, UHMWPE, PP) are produced by metallic catalysts, in an olefin insertion coordination polymerization mechanism (Figure 4). This is one of the largest industrial applications of catalysts. These polymers are mostly produced by Ziegler-Natta classical heterogeneous multisite catalysts; also, HDPE is partly polymerized by Phillips Co. catalyst. There are also growing processes using single site (metallocene) catalysts and during the

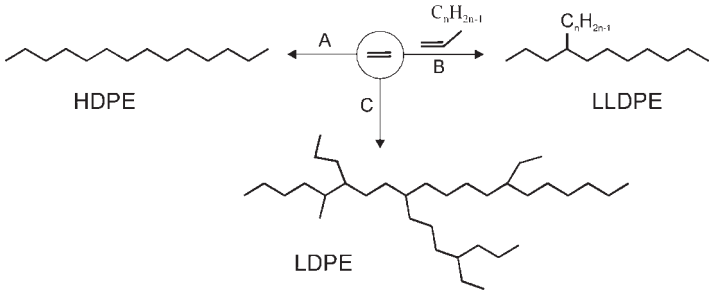


Fig. 4 –. The structure of polyethylene molecules in dependence on polymerization reaction mechanism: A) coordination homopolymerization, B) coordination copolymerization, C) free radical homopolymerization

last two decades an increasing number of polymerization processes employs post-metallocene (or non-metallocene) catalysts.

3.1. Low-density polyethylene

The free radical polymerization processes have the longest history as afore mentioned LDPE was discovered at ICI Co. in 1933 and first production was performed in early 1940s. The free radical polymerization is initiated by oxygen or peroxides and affords production of LDPE with long and short branches. Due to severe conditions, LDPE has an uncontrolled length of branches induced by side reactions. Thus, the so-called "back biting" mechanism forms short chain branches (SCB), usually ethyl, butyl and vinyl groups, and chain transfer to polymer molecule backbone forms long chain branches (LCB). The total number of SCB can range from 15 to 30 per 500 repeated units (1000 C atoms of polyethylene) with a few (2 – 3) LCB. The existence of high degree of SCB and LCB in PE decreases the crystallinity (40 – 60 %) and crystalline melting temperature (105 – 115 °C) influencing its mechanical properties. (Janović, 1997). Polyethylene with appropriate molecular weight can be obtained only under high pressure and temperature resulting from the fact that the ratio of propagation (k_p) and termination (k_t) rate constants $k_p/k_t^{1/2}$ is far too small at ordinary pressures. When the conversion in ethylene polymerization reaches 20 – 30 % it should be terminated due to high viscosity in reactor. The major advantage of this process is allowing for the production of PE film with exceptional clarity having no catalytic residue, needed for many applications, particularly in packaging. Additionally, an advantage is the ability to copolymerize a variety of polar functional monomers in order to produce new polymers such as poly(ethylene-*co*-vinyl acetate) (EVA). Propylene is not sensitive to radical initiation. (Chum et al. 2008)

The polymerization process is performed in a batch reactor (autoclave) or nowadays more frequently in a heavy-walled tubular reactor under 2000 – 3000 bar and 200 – 300 °C. Since the heat of ethylene polymerization is extremely high (3300 kJ kg^{-1}), it is necessary to jacket the reactor with an efficient heat exchange medium. Tubular reactors generally have a number of temperature zones depending on the kind of initiator mixtures used and are operated in nearly steady continuous flow. The product properties are controlled by adjusting key operating variables such as temperature profile across the reactor, the concentration of various peroxide initiators, type of chain transfer agent, and operating frequency of product discharge value. The unreacted monomer is flashed and recycled (Figure 5, Janović, 2011)

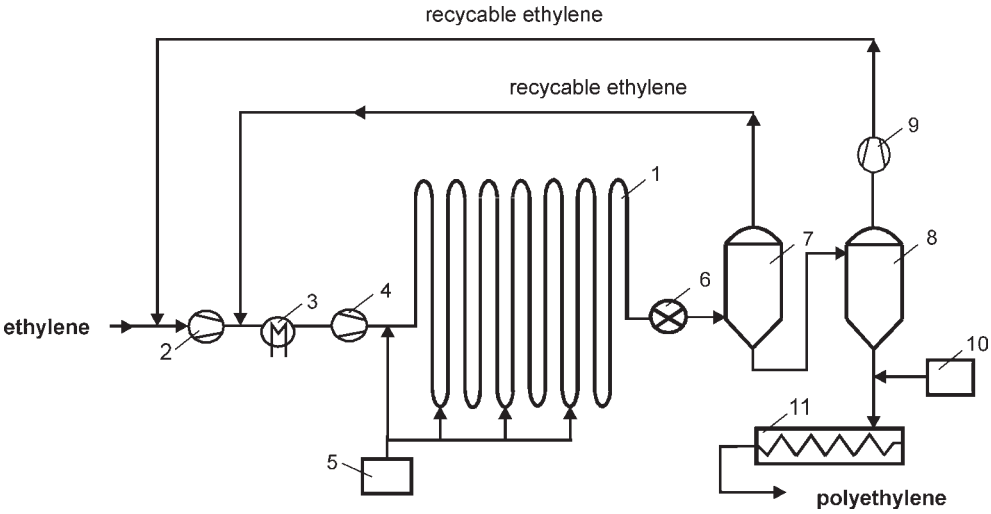


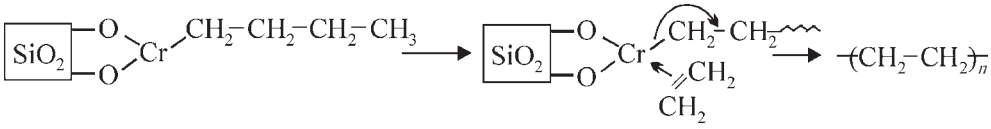
Fig. 5 –. Schematic representation of high-pressure ethylene polymerization at loop reactor: 1 – loop reactor, 2 – compressor, 3 – cooler, 4 – compressor, 5 – dosage of initiator and chain transfer agent, 6 – reduction valve, 7 – high-pressure separator, 8 – low-pressure separator, 9 – low-pressure compressor, 10 – dosage of additives, 11 – extruder

4. Organometallic catalysts

High density polyethylene (HDPE) has a different structure than LDPE. It has a much lower degree of branching (0.5 – 3 / 1000 C) and therefore the marked tendency to crystallization. Due to the fact that transfer to polymer is not possible in coordination polymerization mechanism, long chain branching is highly limited. Thus, HDPE is referred to as linear PE having high degree of crystalline part (70 – 90 %), which increases the density (0.94 – 0.96 g cm^{-3}) and crystalline melting temperature (133 – 138 °C) and determines its mechanical properties. As mentioned before, HDPEs are produced nowadays by three classes of catalysts and processes: Phillips, conventional Z-N and in a less extent, metallocene catalysts.

4.1. Phillips catalysts and processes

Until the 1950s the only PE production was based on free radical, high pressure polymerization processes when P. Hogan and R. Banks at Phillips Petroleum Co. discovered new metallic catalysts based on the chromium-oxide complexes. (Hogan et al., 1986) The catalyst is prepared by adsorption of chromium (VI) oxide (2 – 5 %) onto an amorphous silica support and subsequently by exposure to H_2 and ethylene. In this step the Cr^{6+} is reduced to Cr^{3+} having three orbital places where propagation polymerization reaction takes place by coordination of ethylene molecule:



The Phillips processes produce various types of HDPE with particularly high molecular weight but are not used for LLDPE and are unable for propylene polymerization. The reaction takes place under mild conditions at pressure up to 25 bars and temperature from 85 – 110 °C, usually in a loop-slurry (Phillips) reactor. The traditional slurry polymerization reactors are continuously stirred tank reactors (CSTRS) employing heterogeneous catalysts suspended in solvents like isobutane and are so introduced into the reactor. Because the polymerization of α-olefins is a highly exothermic process the heat generated has to be effectively removed. Therefore, the process is performed in slurry-loop reactor, usually in the Phillips version of double slurry-loop reactor (Figure 6) (Severn et al. 2005). It consists of long circular pipes cooled with a jacket, the formed polymer being driven around the loop by large capacity pump. The polymer is separated by flashing of the volatile diluent which is than recycled.

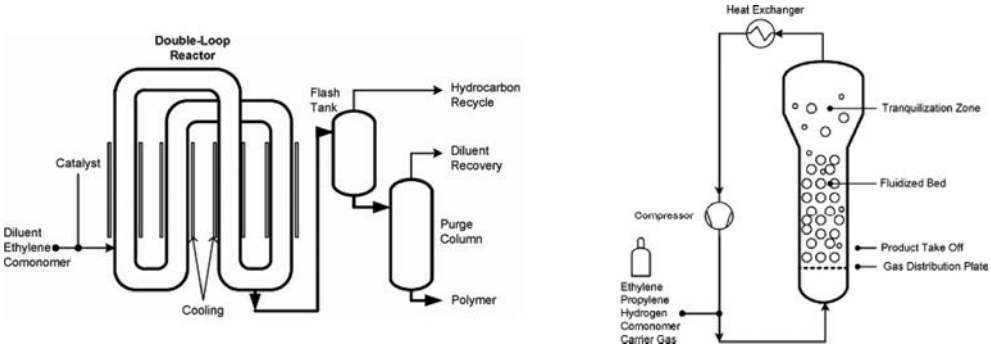


Fig. 6 – Schematic representation of the Phillips (double) slurry loop process (left) and fluidized bed gas-phase polymerization process (right).

4.2. Ziegler Natta catalyst

The discovery of Z-N catalyst gives a new dimension to the polymer science and engineering. The remarkable progress in catalytic olefin polymerization has been achieved by catalyst design, polymer reaction engineering and process technologies which brought about a production of novel and modified polyolefin materials to meet the demands of highly diversified materials.

4.2.1. Conventional catalyst

The organometallic catalysts first discovered in 1953 by K. Ziegler for ethylene polymerization and further developed by G. Natta for stereospecific polymerization of propylene are complexes formed by reaction of transition metal halides such as $TiCl_4$ (but also Zr, V, Mo, Cr, W etc.) and metallic alkyls or alkyl chlorides such as $Al(C_2H_5)_3$, $Al(C_2H_5)_2Cl$, $Al(C_2H_5)Cl_2$, $AlCl_3$ (but also Zn, Mg, Be, Li etc.). Upon mixing the two constituents (Al-alkyl and Ti-halides) in an inert solvent fine crystalline particles are formed, which are very efficient as heterogeneous polymerization catalysts for the monomers of low polarity like α-olefins. Polar compounds (polar solvent, water, oxygen, carbon monoxide and dioxide, carbon sulfide and the like) block the active center of the catalyst and therefore act as catalyst poisons. The "heterogeneous identity" of the catalyst multisided active sites typically results in a broad molecular weight distribution (MWD ≈ 4 – 12) of the polyolefin produced. (Soga et al. 1997; Kissin, 1985)

The polymer grows through alkene, e.g. propylene coordination forming π-complex with the vacant orbital of Ti-atom at the catalyst surface (Figure 7) (Janović et al. 2015). This step is responsible for the versatility of those catalysts as it controls the macromolecular microstructure by the subsequent stereospecific insertion of the alkene into the Ti-C bond.

The titanium complex continues to insert stereospecifically additional propylene monomers between itself and the last group added to the chain, thus repeating the

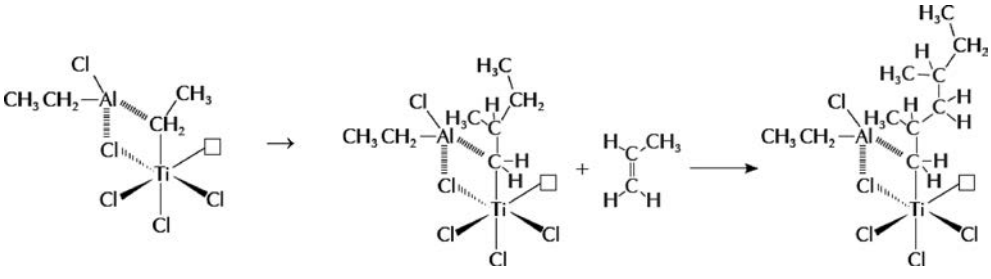


Fig. 7 – Mechanism of propagation reaction in the Ziegler-Natta catalyzed propylene polymerization

reaction and the isotactic polymer is formed. Also, the side group such as in propylene could possibly influence the stereospecificity of the addition to catalyst site. (Janović et al. 2010) Hydrogen is the standard chain transfer agent in olefin polymerization and is used for both ethylene and propylene polymerization processes to lower the molecular weight. Conventional, classical Z-N catalysts for the production of polyolefins have been continuously improved to increase their activity and stereospecificity, as it is shown in Table 1. (adopted from Galli et al. 2001).

Table 1 – Progress in Ziegler-Natta catalysts development for the propylene polymerization

Generation / Year	Catalyst composition	Productivity kg PP/g cat.	Isotactic index[b]	M_w / M_n	Process
1st / 1957 – 1970	$3TiCl_3/AlCl_3 + AlEt_2Cl$	1 – 2	88 – 91	-	solution / slurry
2nd / 1970 – 1978	α-$TiCl_3 + AlEt_2Cl$	3 – 5	95	-	solution / slurry
3rd / 1978 – 1980	$TiCl_4$/ester/ $MgCl_2$ + $AlEt_3$/ester[c]	5 – 15	98	8 – 10	Liquid propylene
4th / 1980	$TiCl_4$/diester/ $MgCl_2$[d] + $AlEt_3$/silane	20 – 60	99	6.5 – 8	RGT[e]
5th / 1999	$TiCl_4$/dieter/ $MgCl_2$ + $AlEt_3$	50 – 120 (max. 2400)	99	5 – 5.5	RGT[e]
6th / >2000	metallocene (Zr) + MAO	5×10^3	99	2 – 3	solution
7th / 2010	multicatalyst: Z-N + metallocene	9×10^3	99	2 – 3	gas phase

[a] polymerization conditions: liquid propylene, 70 °C, H_2;
[b] extracted part of iPP;
[c] esters: alkyl benzoates, silanes (dialkyl dialkoxy silanes);
[d] phtalates;
[e] RGT – Reactor Granule Technology

The two most important development steps in Z-N catalysts were: (1) the use of $MgCl_2$ as a reactive catalyst carrier of large surface due to the similarities in the crystal structure and ionic radii with $TiCl_3/TiCl_4$, which highly improve the polymerization yield and (2) discovery of the selective Lewis bases as electron donors, such as alkyl benzoates, phtalates, succinates, and silanes (Andoni et al., 2008), which significantly improve the catalyst stereospecificity (Figure 8; Janović, 2015). The resulting catalyst displays an activity from about 5 kg PP / g Ti to as high as 100 – 1000 kg PP / g Ti (Table 1). Also, porous structure of catalyst particles allowed for the control of polymer morphology such as in the so-called spherical process. Regarding to their high activity, the residual amounts of catalysts were low and catalyst removal steps were no longer performed as a part of the production process.

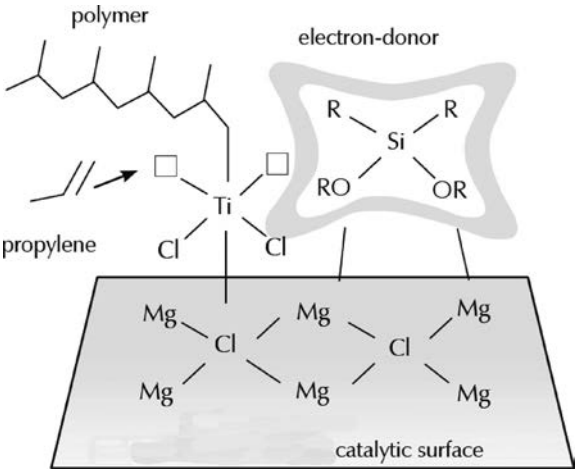
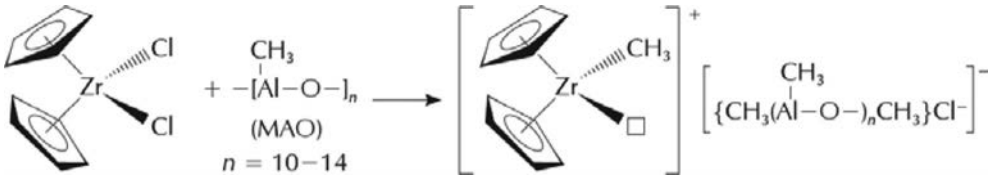


Fig. 8 – Reaction scheme of α-olefins polymerization by Ziegler-Natta catalyst on $MgCl_2$ carrier.

4.2.2. Metallocene catalysts

As it was mentioned, the huge progress in Z-N polymerization was made by discovery of homogeneous metallocene catalysts, complexes formed by organic ligands like ferrocene connected via π-bonds to transition metal atoms such as Ti, Zr, Hf, Sc, Th and others. Besides cyclopentadienyl, the often used organic ligands are indenyl and fluorenyl and their substituents. Very high activities were achieved only by using cocatalysts, particularly methyl aluminoxane (MAO) (Kaminsky, 2012), whereat single site active centers are formed as metallocenium cations:



The presence of MAO here boosted the activity of the metallocene based catalysts and produced uniform polyethylene with a narrow molecular weight distribution. The monomer is coordinated with a highly electrophilic cationic complex, followed by insertion of the monomer unit in the metal-carbon bond to produce a growing polymer chain (Figure 9; Janović, 2015).

A number of metallocene catalysts is known. They differ by the kind of ligands or metals and therefore by using them it was possible to obtain polymers of different stereoregularity, controlled molecular weight, short and long chain branching, and

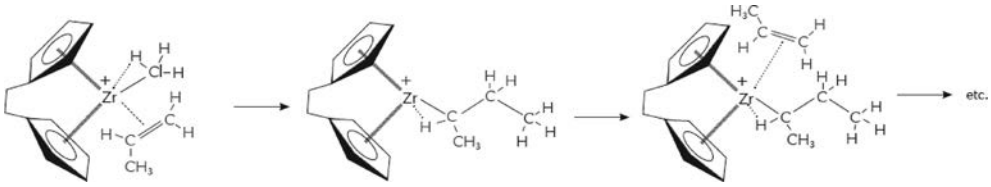


Fig. 9 – Mechanism of propagation reaction in the metallocene catalyzed propylene polymerization

comonomer distribution in copolymers. As a function of catalyst structure and process conditions it is therefore possible to tailor polyolefin materials according to their particular application. (Schiers et al., 2000)

The most attractive point of metallocene catalysts is the easy tailoring of microstructure of obtained polyolefin by convenient substitution of the cyclopentadienyl group(s) with larger aromatics such as fluorenyl or benzindenyl and bridging them with silylene moieties commonly known as *ansa*-metallocene. (Brintzinger et al. 2013) Among thermoplastic polymers, polypropylene is quite outstanding with respect to its versatility regarding different tacticity; atactic, isotactic, syndiotactic, hemitactic, and stereoblock structures are obtained depending on the ligand structure as it is shown in Figure 10 (adopted from Mülhaupt, 2003).

It is interesting to mention the stereoblock polypropylene(s) containing alternating flexible and crystalline isotactic segments. When the molecular weight is high

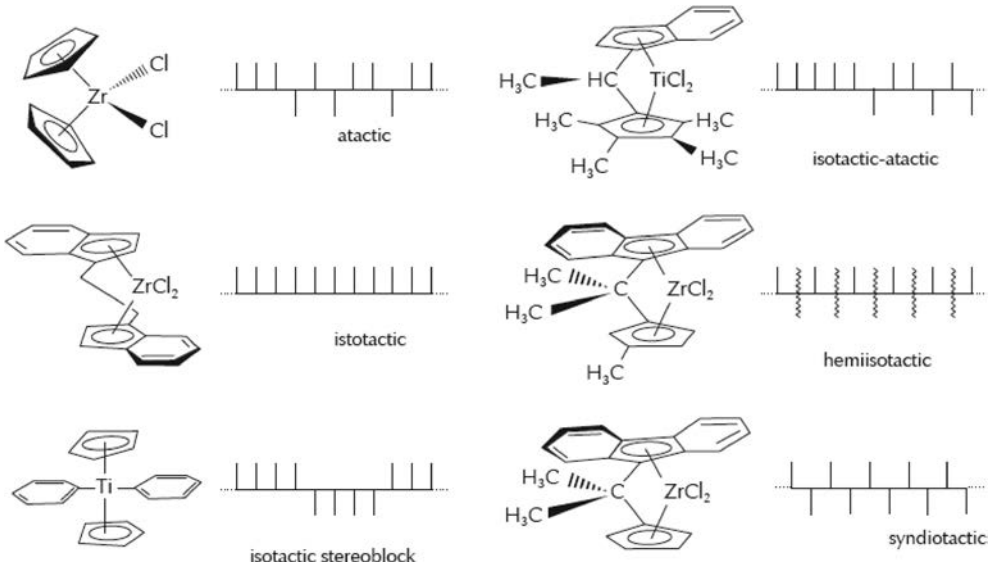


Fig. 10 – Influence of metallocene catalyst structure on propylene polymerization and polypropylene stereoregularity

enough, then more than two isotactic segments may be present that are sufficiently large to cause thermally reversible cocrystallization and behavior of the thermoplastic elastomers. The performance of metallocene-based catalysts systems was improved, it became possible to produce isotactic, syndiotactic and stereoblock PP on an industrial scale. Also, ethylene/1-olefin copolymers with high 1-olefin content, cycloolefin copolymers, ethylene/styrene copolymers, syndiotactic polystyrene, and long chain ethylene copolymers became available.

In the last 20 years the field of metallocene catalysts is extended to new metallic complexes of Ni, Pd, Co, and Fe and bearing non-cyclopentadienyl ligands. They are named "non-metallocene" or "post-metallocene" single-site catalysts and they require an activator (cocatalyst) to generate the catalytically active center (Figure 11, Baier et al. 2014). Highly active Ni, Pd and Co complexes with diimine ligands are capable of polymerizing ethylene to high molecular weight polyethylene having short chain branches, in the absence of α-olefin comonomers.

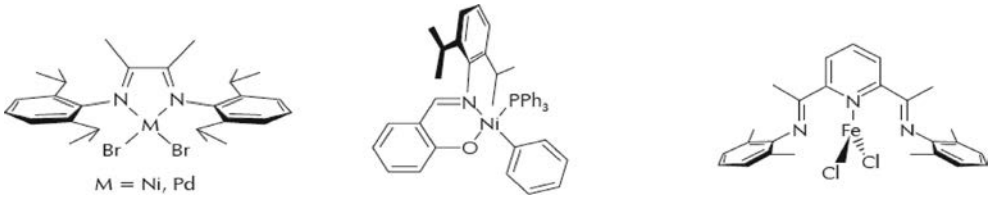


Fig. 11 – Structure of some important post-metallocene catalysts.

The Pd catalyst provides high molecular weight PE, a highly branched amorphous material with about 100 methyl, ethyl, propyl and higher branches per 1000 C atoms. Remarkably, these catalysts can produce polar polyolefins by copolymerization of α-olefin with polar monomers, preferably acrylates and methacrylates due to lower oxophilicity of these catalysts, compared to metallocenes. (Nomura et al., 2008) The structurally related iron complexes with bulky bis-imine ligands were remarkably active in ethylene polymerization giving linear high density PE. These catalysts polymerize ethylene to a strictly linear polymer rapidly, however iron pyridyl-diimine complexes with less bulky N-aryl substituents (R = H in Figure 11.) are very active catalysts for ethylene oligomerization to linear higher oligomeric 1-olefins. Substantially linear PE with typically a few LCB per 1000 C of polymer molecules is made by postmetallocene catalysts (Pritchard, 2012) in a reaction of terminal branching.The breakthroughs in ethylene homo- and copolymerization by means of single site catalysts stimulated the industrial development of high density, linear low density, and very low density polyethylenes and a new ethylene / 1-octene thermoplastic elastomer. (Baier et al., 2014)

5. Linear low density polyethylene

Linear low density polyethylene (LLDPE) is similar in properties to LDPE, however, it is one of the mechanically most flexible materials among polyolefins and accordingly has found the appropriate applications. The short chain branches (SCB) are formed by copolymerization of ethylene and comonomers, smaller amount (up to 10 %) of higher α-olefins such as 1-butene, 1-hexene, 1-octene and other α-olefins which determine the precise length of branches (Figure 4.). LLDPE is produced only by coordinating catalysts, predominantly by Z-N systems using slurry or gas phase processes under mild process conditions (Figure 6). Physical properties such as density and melting point can be tuned by the type and concentration of the comonomers and by polymerization conditions. By increasing the amount of higher α-olefins, the VLDPE could be also produced.

In the last several years, a relatively high number of single-site metallocene metal complexes have emerged as easily tunable catalysts for the production of tailored LLDPE materials. They present a higher efficiency for the comonomer incorporation, and therefore, the lower feeding of the comonomer is needed. Also, VLDPE could be obtained by using metallocene catalysts having some advantages such as easier comonomer incorporation. By using α-diimino nickel complex (Figure 11), M. Brookhart with collaborators (Johnson et. al. 1995) developed the synthesis of branched polyethylene using only ethylene, as a single feed, where a part of ethylene was transferred into 1-butene that formed the copolymer.

A relatively new approach for the LLDPE syntheses is the use of a multiple catalyst system capable of simultaneous ethylene dimerization and copolymerization of

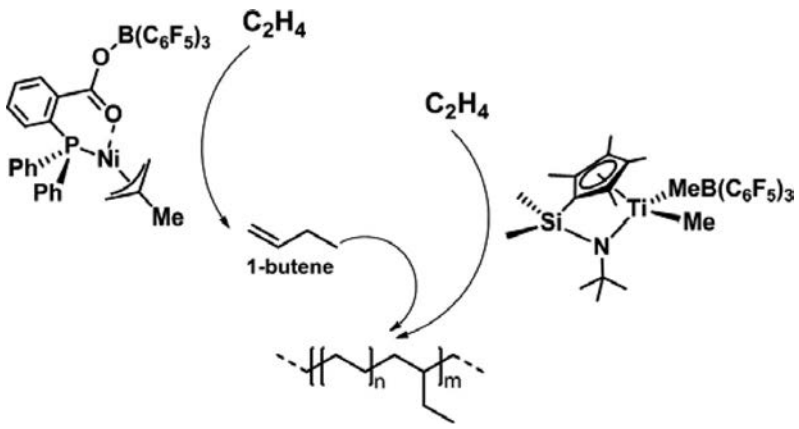


Fig. 12 – Synthesis of LLDPE via tandem catalysis of metallocene Ni and Ti complexes using ethylene as a single monomer feed.

1-butene formed *in situ* with ethylene. In both cases, the use of relative expensive and very pure higher α-olefines as co-monomers is avoided (Musikabhumma et al. 2003).

Similarly, LLDPE from only ethylene in the single reactor may be produced by a tandem catalysis systems consisting of two metallocenes: catalyst-1 trimerizes ethylene to 1-hexene and catalyst-2 copolymerizes 1-hexene with ethylene. (Zhang et al., 2007) However, these smart synthetic routes and following processes are not yet cost effective.

6. Bimodal polyethylene

Recently, a trend towards the production of polyolefins with advanced properties by in-reactor blending (whether polymer-polymer compositions or nanocomposites) and multi-stage processes has been followed. Basically all known polymerization processes can be operated in a series of reactors in order to achieve multimodal PE. To get polymers with broad to very broad molecular weight distribution and to assure an acceptable processing, the reactors are run in a cascade and at very different process conditions. (Ruff et al., 2012) In each polymerization step a variation of the polymer in morphology, molecular weight and chemical composition is obtained. Therefore, such PEs of medium density grades exhibit the strength and stiffness of HDPE while retaining the high stress crack resistance, and as a main point, the high resistance to the slow crack propagation under stress. They are mostly being applied for constructing pipes for drinking water, sewage and natural water. They withstand a minimum circumferential stress of 10 MPa for 50 years at 20 °C.

Presently, there are two bimodal PE processes depending on the catalyst used: (a) Z-N/Mg supported Ti catalyst and (b) metallocene catalyst. Using Z-N catalyst, molecular weight is adjusted with addition of hydrogen over a wide range without a loss in activity. In the first reactor high amounts of hydrogen are fed with ethylene. The process leads to the formation of PE of low molecular weight. The second reactor is loaded with much less hydrogen to form PE with broad or very broad molecular weight distribution. (Böhm, 2003)

The cascade process allows for a further essential product modification: the incorporation of comonomer in the long polymer chains within the second reactor. In this way a polymer blend is formed with a low-molecular-weight ethylene homopolymer and a high-molecular-weight ethylene-1-alkene copolymer. These products build up a polymer alloy in the solid state consisting of crystalline regions

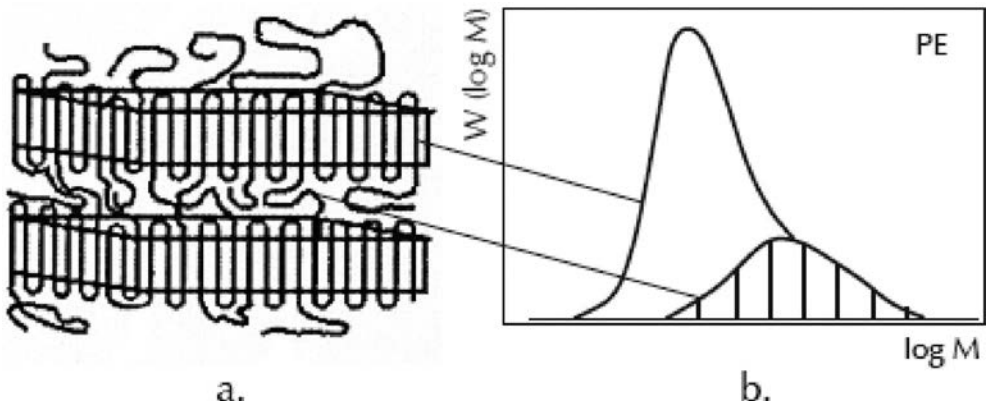


Fig. 13 – Structure of bimodal polyethylene: morphology (a), molecular weight distribution (b).

(crystal lamellae) and amorphous regions in between. The crystalline regions are mainly formed by the low-molecular-weight homo-polyethylene. The high-molecular-weight copolymers form the amorphous regions and act as tie molecules that connect the crystal lamellae. In this way, a physical network is formed (Figure 13; Böhm, 2003). By changing the ratio of these phases the polymer properties will change as well.

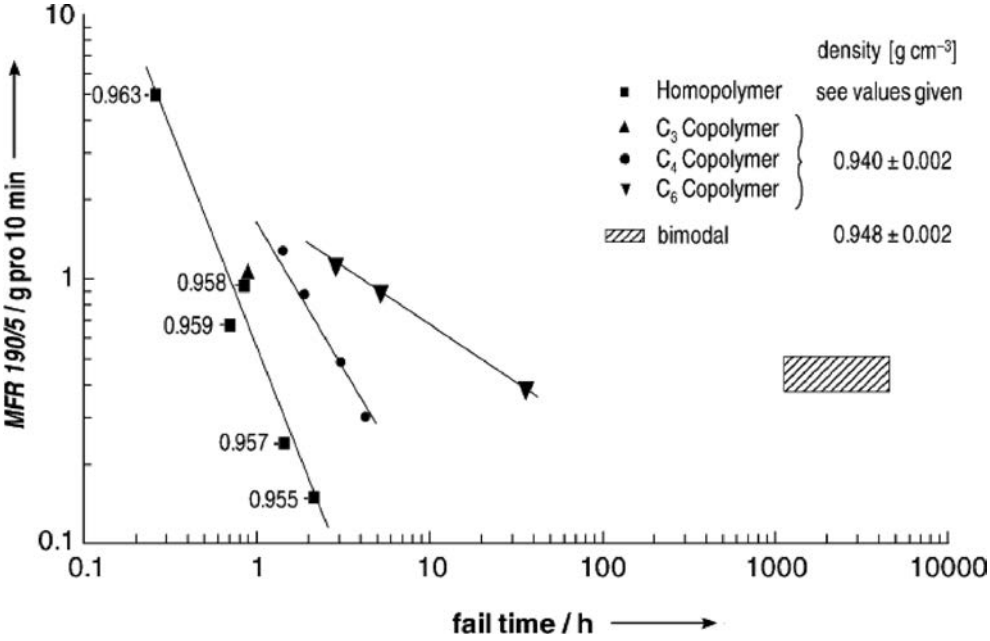


Fig. 14 – Fail-time test and product data for polymer blends.

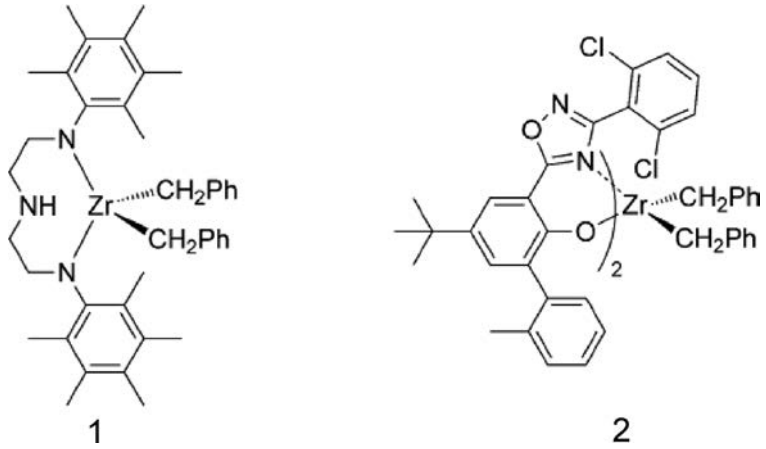


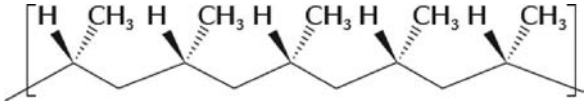
Fig. 15 – Structure of Zr-metallocene diamido-amine (1) and oxadiazol (2) catalysts for bimodal polyethylene production

The comparison of some important properties of polyethylene and its higher α-olefin copolymers with bimodal polyethylene (such as the melt flow rate, MFR, and fail time) is presented in Figure 14., Boehm 2003).

The metallocene compound for the production of bimodal PE consists of the two very active catalyst precursors immobilized together on a support, such as silica; ethylene is polymerized in the presence of hydrogen in a single fluidized bed gas phase reactor (Baier, 2014).The Zr-metallocenes, the diimido-amine complex (1) with a high H_2 response produces the low molecular mass component of the blend which can be adjusted by hydrogen partial pressure(ca. $5 – 35\times10^3$ g mol^{-1}) and oxadiazol complex (2) which exhibits essentially no H_2 response will produce high molar mass fraction (ca. $4 – 7\times10^5$ g mol^{-1}) (Figure 15, Baier, 2014).

7. Polypropylene

Polypropylene, due to the asymmetry of propylene monomer, exists in three stereochemical configurations: isotactic, syndiotactic and atactic. In isotactic PP the methyl groups are placed on the same side of the backbone chain:



in syndiotactic PP they are placed on the regularly alternating sides; and in the atactic form the methyl groups are arranged randomly along the chain.

Both isotactic PP (iPP) and syndiotactic PP (sPP) are semicrystalline polymers, whereas atactic form is amorphous and it has a little commercial value. iPP forms a helical conformation, which easily crystallizes (60 – 70 %) and has a high crystalline melting temperature (174 °C). The wide range of iPP applications as well as its simple production with Z-N catalysts resulted with its huge commercial success. Therefore, almost all industrial processes use these catalysts and in less extent also some metallocenes due to the combination of outstanding physical and mechanical properties, low density and ease processability and recycling. As it was mentioned earlier, since the first discovery of PP by G. Natta in 1954, a wide variety of different catalysts has been designed and developed (Table 1), leading to tailored polymers of entirely different structures (Figure 10) and applications by controlling polymer tacticity, molecular weight and its distribution. The structure and physical properties of sPP have been described elsewhere. (De Rossa at al. 2006) One of shortcomings of iPP's properties is its limited impact resistance, particularly at low temperatures. Enhancement of the impact properties could be achieved by dispersing various types of elastomers, usually ethylene – propylene copolymer rubber (EPR) in a volume fraction of 25 – 30 %, within iPP matrix. However, the two phases formed is mutually highly incompatible, which is the main drawback of this process. (Grein et al. 2003) An alternative for blending process is the preparation of random copolymers of propylene with ethylene, 1-butene or 1-hexene. For propylene-ethylene copolymers, the values of elastic modulus, ductility, toughness and strength can be tuned by changing the share of the ethylene. Such copolymers are usually produced using two reactors in series with Z-N catalyst or supported metallocenes. (Gahleitner et al., 2013) The first reactor makes iPP and transfers it into the second reactor where propylene-co-ethylene polymer is produced. The copolymer component is amorphous and intimately dispersed within the homopolymer phase, like in bimodal polyethylene (Figure 13), even though the two phases are immiscible. The copolymer phase dissipates energy during impact, thus increasing the impact resistance at low temperatures.

Modern bulk (liquid monomer) and gas-phase processes are preferable in the production of PP homopolymer, random copolymers and high impact strength copolymers. The most widely adopted process for PP is Spheripol (P. Galli and coworkers at Montedison, Italy), the so called reactor granule technology (RGT). (Galli et al., 2001) A leading advantage of the Spheripol process, in combination with the appropriate catalyst, is its unique ability to produce polymer spheres directly in the reactor. The process is based on the use of Z-N/$MgCl_2$ supported, Lewis acid-based catalyst in a form of spherical grains composed of agglomerated porous micro particles (Figure 16, adopted from Polymirae, 2015).

The process consists of three main steps: catalyst and raw material feeding, polymerization reaction and finishing. The catalyst, liquid propylene and hydrogen for molar mass regulation are continuously fed into the loop reactor where bulk polym-

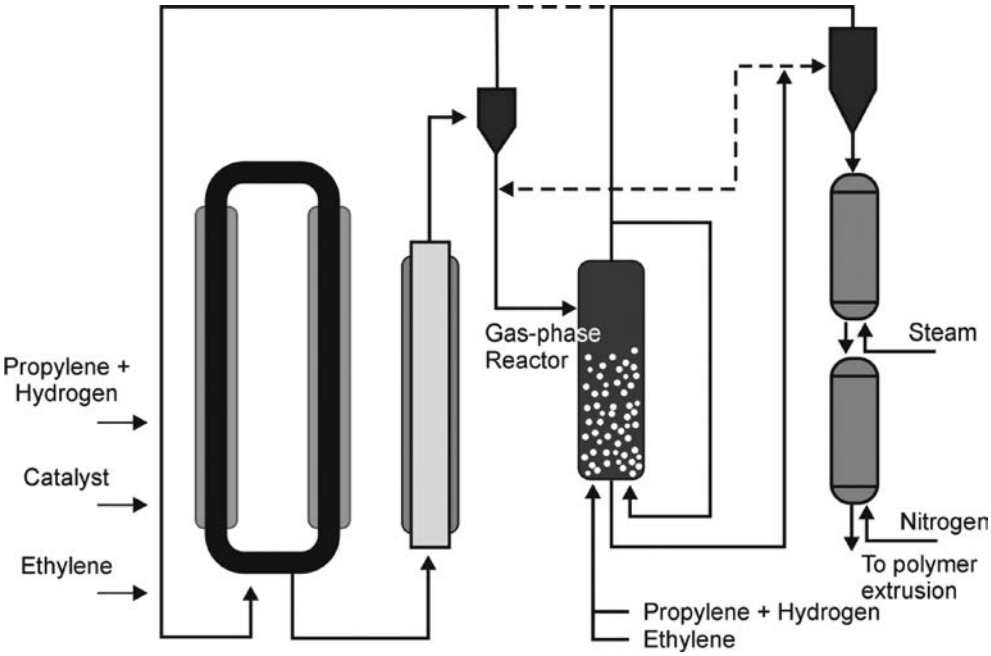


Fig. 16 – Spheripol process: prepolymerization combined with liquid phase propylene polymerization in the loop reactor and subsequent ethylene/propylene gas phase polymerization.

erization occurs; the option is subsequent gas phase copolymerization reaction. The finishing section consists of the separation of polypropylene and unconverted monomers and their recycling back to the reactor.

In the prepolymerization reaction the agglomeration is achieved; thus formed catalytic particles serve as a template for the formation of iPP particles with controlled porosity in the first polymerization stage in slurry-loop reactor, in liquid propylene. Since the deagglomerated catalysts remain active, it is possible to polymerize other olefins in the pores of spherical PP grains in a fluidized bed, gas phase reactor to obtain a number of desirable polymer blends, such as impact-modified PP blends with ethylene-propylene rubber. (Figure 17; Mülhaupt, 2003)

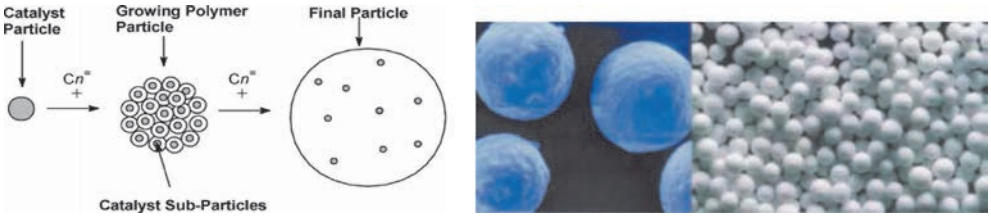


Fig. 17 – Polymer particle growth in propylene polymerization with Z-N microporous catalysts.

These outstanding properties are not easily accessible with propylene-based copolymers produced with conventional Ziegler-Natta catalysts because of nonrandom distribution of comonomers and non-homogeneous composition. However, the much higher compatibility has been obtained by using appropriate metallocene catalysts. (De Rossa et al., 2007)

8. Polyolefin nanocomposites

Polyolefin nanocomposites are polymeric materials with novel properties, such as improved mechanical properties like stiffness, strength or toughness, increased heat distortion temperature, reduced permeability and flammability and improved electrical and optical properties. Therefore, polymeric nanocomposites are used as replacement for materials like metals, glass and ceramics. The filler used should fulfill relevant characteristics like chemical surface composition, size and shape of particles, structure and pore size, hydrophobicity, mechanical and thermal properties. The properties of the nanocomposites are not only influenced by the kind of fillers but also by the microstructure of the polyolefins and the preparation process. Metal powders, metal oxides, carbonates, silica, talcum, layered silica as well as fibers such as carbon nanotubes, carbon nanofibers and polymer fibers are used as fillers. The use of layered inorganic materials such as multilayered silicates, clay, montmorillonite or graphene is of special interest, because of the additional great barrier resistance against permeation of gases and liquids. (dos Ouros et al., 2014)

The use of nanofillers in polymeric composite materials has a marked advantage in comparison to the microfillers. Nanofillers will improve the mentioned properties of material in even small content without a serious influence on the increase of its density. The influence of the filler particle dimensions (the ratio between length (L) and diameter (D)) on the flexural modulus of polypropylene composites with different filler content is shown in Figure 18. (Janović, 2012)

Today, most polymer composites are prepared by mechanical blending of the particles or fibers above the melting temperature of the desired matrix. Melt compounding of polyolefins with nanoparticles often leads to an insufficient filler dispersion, especially at a high filler content, which leads to aggregation, which in turn deteriorates the mechanical properties.

Such disadvantages can be solved by “in-situ” polymerization whereby the cocatalyst methylaluminoxane (MAO) can be adsorbed or anchored on the surface of the reinforcing nanoparticles, such as fibers, carbon nanofibers (CNF), multi-walled carbon nanotubes (MWCNT), thereby changing the surface to a hydrophobic one.

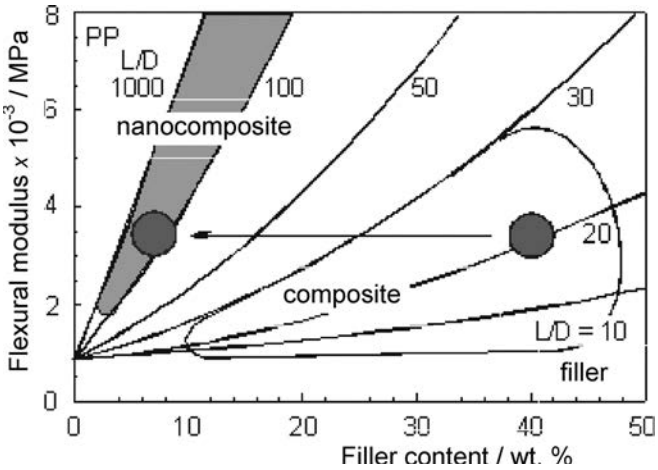
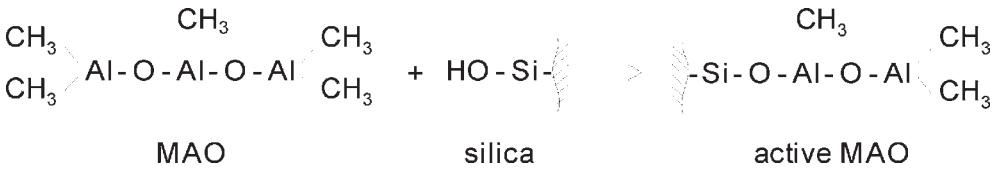


Fig. 18 – Flexural modulus of PP composites in dependence on content and size of filler particles

An efficient technique to immobilize the MAO at the surface of the filler is the "polymerization-filling technique". (Kaminski, 2014). In the first step MAO reacts, for example, with the OH-groups of silica or with carboxyl groups of oxidized carbon nanotubes or it is physically adsorbed at the surface:



In the second step, metallocene with the desired properties is added, forming catalytically active polymerization sites on the (nano)surface. The thickness of the polymer film, formed by the addition of ethylene or propylene, depends on the polymerization conditions, especially the polymerization time, the kind of metallocene catalyst, and the pressure of the monomer. The in-situ polymerization leads to composite materials where each particle or fiber is intensively covered with the polymer giving far better particle dispersion for reasonable amounts of nanofillers in the final polymer (0.5 – 2.0 %).

Heterogeneous nanosized Ziegler-Natta catalysts can also be used for in-situ polymerization of layered materials. Low nanoparticle contents may be sufficient to obtain new or modified material characteristics, particularly a faster crystallization rate and a higher crystallization temperature. In particular, polyethylene and isotactic or syndiotactic polypropylene nanocomposites have been investigated and required as materials for electronic and magnetic sensors (Coiai et al. 2015), domes-

tic appliances and automotive applications with outstanding properties. Carbon nanofibers (CNF) or multi walled carbon nanotubes (MWCNT) (McNally et al. 2011) are especially attractive classes of fillers for polymers because of their intriguing mechanical and thermal properties. Inorganic particles, especially silica balls for polyolefin nanocomposites and carbon nanotubes are used to increase the stiffness to improve the form stability and other properties of polyolefins.

9. Conclusion

Polymeric materials belong to the most important classes of the presently used technical materials. They have made tremendous progress in the last several decades. Among them, polyolefins based predominantly on ethylene and propylene homopolymers, copolymers, blends and composites, account for more than half of the synthetic polymers production of about 300 million tons yearly. The progress started by the discoveries of organometallic catalysts and stereospecific polymerization of α-olefins by Nobel Prize awarded scientists Karl Ziegler and Giulio Natta and has led to outstanding and continuous development of both catalytic polymerization reactions and manufacturing processes. These discoveries have stimulated an intensive basic and applied research all over the world, leading to the advanced classes of metallocene catalysts and great development of polymer production and demand. Advanced catalyst systems, the key of modern polymerization processes, are available to produce polyolefins with a wide range of structures, properties and applications. According to P. Galli (G. Natta Institute, Ferrara, Italy) polyolefins are the most promising large-volume materials for the 21st century.

Acknowledgment

The authors would like to thank Zorica Veksli (Rugjer Boskovic Institute, Zagreb) for her contribution in literature selection, valuable discussion and significant improvement of the overall manuscript.

References

[1] A. Andoni, J. C. Chadwick, H. J. W. Niemantsverdriet, P. C. Thüne, The role of electron donors on lateral surfaces of $MgCl_2$-supported Ziegler-Natta catalysts: Observation by AFM and SEM, J. Catal. 257 (2008) 81-86.

[2] M. C. Baier, M. A. Zuideveld, S. Mecking, Post-metallocene in the industrial production of polyolefins, Angew. Chem. Int. Ed. 53 (2014) 9722-9744.

[3] L. L. Böhm, The ethylene polymerization with Ziegler catalysts: fifty years after discovery, Angew. Chem. Int. Ed. 42 (2003) 5010-5030.

[4] H. Brintzinger, D. Fischer, Development of *ansa*-metallocene catalysts for isotactic olefin polymerization, Adv. Polym. Sci. 258 (2013) 29-42.

[5] V. Busico, Giulio Natta and the development of stereoselective propene polymerization, Adv. Polym. Sci. 257 (2013) 37-58.

[6] S. Coiai, E. Passaglia, A. Pucci, G. Ruggeri, Nanocomposites based on thermoplastic polymers and functional nanofiller for sensor applications, Materials 8(6) (2015) 3377-3427.

[7] P. S. Chum, K. W. Swogger, Olefin polymer technologies – History of recent progress at the Dow Chemical Company, Prog. Polym. Sci., 33 (2008) 797-819.

[8] M. Gahleitner, L. Resconi, P. Doshev, Heterogeneous Ziegler-Natta, metallocene, and post-metallocene catalysis: Successes and challenges in industrial application, MRS Bulletin, 38(3) (2013) 229-233.

[9] P. Galli, G. Vecellio, Technology: driving force behind innovation and growth of polyolefins, Prog. Polym. Sci. 26 (2001) 1287-1336.

[10] C. Grein, K. Bernreitner, A. Hauer, M. Gahleitner, W. Neißl, Impact modified isotatic polypropylene with controlled rubber intrinsic viscosities: Some new aspects about morphology and fracture, J. Appl. Polym. Sci. 87 (2003) 1702-1712.

[11] J. P. Hogan, R. L. Banks, History of crystalline polypropylene in R. B. Seymour, T. Cheng (eds.), History of Polyolefins, D. Reidel Publishing, Co., Dordrecht, The Netherlands, 1986. p. 103-115.

[12] Z. Janović, Z. Veksli, Šezdesetgodišnjica Ziegler-Nattinih katalizatora i stereospecifičnih polimerizacija, Kem. Ind. 64 (2015) 363-379.

[13] Z.Janović, Petrokemija-razvitak i mogućnosti, (engl. Petrochemistry-Development and prospective) HAZU, Bilten raz.teh. znanosti 12 (2012) 7-26.

[14] Z. Janović, Naftni i petrokemijski procesi i proizvodi (engl. Petroleum and petrochemical processes and products), 2nd Ed., Hrvatsko društvo za goriva i maziva, Zagreb, 2011, p. 308.

[15] Z.Janović, Polimerizacije i polimeri (engl. Polymerization and Polymers), Kemija u industriji, Zagreb, 1997, p.317-329.

[16] ,Z. Janović, A. Jukić, O. Vogl, Spacer groups in macromolecular structures, Polimeri 31(1) (2010) 14-21.

[17] L. K. Johnson, C. M. Killian, M. Brookhart, New Pd(II)- and Ni(II)-Based Catalysts for Polymerization of Ethylene and α-Olefins, J. Am. Chem. Soc. 117 (1995) 6414–6415

[18] W. Kaminsky, Discovery of methylaluminoxane as cocatalyst for olefin polymerization, Macromolecules, 45 (2012) 3289-3297.

[19] W. Kaminsky, Metallocene based polyolefin nanocomposites, Materials 7 (2014) 1995-2013.

[20] Y. V. Kissin, Isospecific Polymerization of Olefins With Heterogeneous Ziegler-Natta Catalysts, Springer Verlag, New York, 1985.

[21] T. McNally, P. Poetschke, Polymer-Carbon Nanotube Composites: Prepration, Properties and Applications, Woodhead Publ., Cambridge, 2011.

[22] R. Mülhaupt, Catalytic polymerization and post polymerization catalysts fifty years after the discovery of Ziegler's catalysts, Macromol. Chem. Phys. 204 (2003) 289-327.

[23] K. Musikabhumma, T. P. Spaniol, J. Okuda, Synthesis of branched poyethylenes by the tandem catalysis of silica supported linked cyclopentadienyl-amido titanium catalysts and a homogeneous dibromo nickel catalyst having pyridyl imine ligand, J. Polym. Sci., Polymer Chem. Ed. 41 (2003) 528-544.

[24] K. Nomura, B. Kitiyanan, Recent progress in precise synthesis of polyolefins containing polar functionalities by transition metal catalysis, Current Org. Synt. 5 (2008) 217-226.

[25] A. C. dos Ouros, M. O. de Souza, H. O. Pastore, Metallocene supported on inorganic solid supports: an unfinished history, J. Braz. Chem. Soc., 25 (2014) 2164-2185.

[26] J. Pritchard, Introduction to polyolefins, in J. B. P. Soares, T. F. L. McKenna (eds.), Polyolefin Reaction Engineering, Wiley-VCH Verlag, 2012, 1-12.

[27] www.polymirae.com/en/product/popup_ spheripol.php

[28] De Rossa, F. Auriemma, Structure and physical properties of syndiotactic polypropylene: A highly crystalline thermoplastic elastomer, Progr. Polym. Sci. 31 (2006) 145-237.

[29] C. De Rossa, F. Auriemma, O. Ruiz de Ballesteros, L. Resconi, I. Camurati, Tailoring the physical properties of isotactic polypropylene through incorporation of comonomers and the precise control of stereo-and-regioregularity by metallocene catalyst, Chem. Mater. 19 (2007) 5122-5130.

[30] M. Ruff, C. Paulik, Controlling polyolefin properties by in-reactor blending, 1 – Polymerization process, precise kinetics, and molecular properties of UHMW-PE polymer, Macromol. React. Eng., 6 (2012) 302-317.

[31] J. Schiers, W. Kaminsky (Eds.), Metallocene-Based Polyolefins, J. Wiley and Sons, New York, 2000.

[32] J. R. Severn, J. C. Chadwick, R. Duchateau, N. Friederichs, "Bound but not gagged" – Immobilizing single-site α-olefin polymerization catalysts, Chem. Rev. 105 (2005) 4073-4147.

[33] A. Shamiri, M. H. Chakrabarti, S. Jahan, M. A. Hussain, W. Kaminsky, P. V. Aravind, W. A. Yehye, The influence of Ziegler-Natta and metallocene catalysts on polyolefin structure, properties and processing ability, Materials 7 (2014) 5069-5108.

[34] K. Soga, T. Shiono, Ziegler-Natta catalysts for olefin polymerization, Prog. Polym. Sci. 22 (1997) 1503-1546.

[35] K. S. Whiteley, T. G. Heggs, H. Koch, R. L. Mawer, W. Immel, Polyolefins, in Ullman's Encyclopedia of Industrial Chemistry, Wiley-VCH, Weinheim, 2005.

[36] J. Zhang, B.-G. Li, H. Fan, S. Zhu, Synthesis of ethylene – 1-hexene copolymers from ethylene stock by tandem action of bis(2-dodecylsulfanyl-ethyl) amine-$CrCl_3$ and $Et(Ind)_2ZrCl_2$, J. Polym. Sci., Part A: Polym. Chem. 45 (2007) 3562-3569.

Review paper

Comparison of PDE-based Sparse Image Inpainting Methods

Irena Galić[1,*], Branka Zovko-Cihlar[2,*]

[1]Faculty of Electrical Engineering, Kneza Trpimira 2B, Osijek, Croatia

[2]Faculty of Electrical Engineering and Computing, Unska 3, Zagreb, Croatia

In recent times image processing and computer vision have become a very attractive fields for applications of partial differential equations (PDEs) and corresponding variational techniques. While applications of PDEs are often associated with denoising problems, there is an increasing number of publications where their potential as inpainting methods is explored. PDEs are frequently used for interpolation problems where most data is given and only filling in the missing data in corrupted areas is needed. It is less frequently used when data is sparser. Goal of the presented paper is to explore the potential of PDEs for a scattered data interpolation problem and importance of selecting good interpolation points for a sparse image or selecting appropriate PDEs if a sparse image is less favourable sparsified. PDE-based interpolation method is described and a number of different PDEs are evaluated on various sparse images.

Key words:

inpainting, partial differential equation (PDE), diffusion, interpolation, sparse image

1. Introduction

In general, PDE-based image inpainting methods were invented to recover some relatively small missing regions by smoothly propagating information from the known surrounding areas. Additional assumptions on the data have to be involved, e.g. smoothness, band limitation. The filling-in effect has also become the main

[*]irena@etfos.hr, Branka.Zovko-Cihlar@fer.hr

feature of PDE-based inpainting methods such as [1]–[6]. Here most of the data is given and one aims at restoring missing information in certain corrupted image areas by means of second or higher-order PDEs. The basic idea is to regard the given image data as Dirichlet boundary conditions, and interpolate the data in the inpainting regions by solving appropriate boundary value problems.

Variational and PDE methods that have been proposed for inpainting have also been investigated for more classical interpolation problems such as zooming into an image by increasing its resolution [7]–[13]. More recently, the variational and PDE-based interpolation and inpainting techniques have been used for scattered data interpolation, although, the sparsity of the scattered data constitutes a real challenge for some of these techniques. While second-order PDEs may satisfy a maximum-minimum principle, they often create singularities at isolated interpolation points in 2-D. Higher-order PDEs, on the other hand, may give smoother solutions but the violation of an extremum principle can lead to undesirable over- and undershoots; see e.g. [10].

Sometimes PDE-based interpolation strategies have been tailored to specific data sets such as surface data in digital elevation maps [14]–[16]. More general discussions on how to reconstruct an image from a suitable set of feature points and their derivatives have been presented by Lillholm et al. [17].

In recent years PDE-based image compression methods are developed where PDE-based scattered data interpolation is applied [18]–[26]. The idea is to keep only a small fraction of pixels to be stored according to certain algorithm and to reconstruct the remaining data with PDE-based interpolation. Regarding the subdivision strategy, many related variable block size image coding algorithms exist, in particular methods based on quadtree decompositions; see e.g. [27] and [28]. Interesting adaptive triangulation ideas can be found in [29], [30] and [31]. In R-EED [23] adaptive triangulation from [21] is replaced by a subdivision into a rectangular structure with different amendments.

The goal of this paper is to explore interpolation qualities of the different PDEs for different sparse sets of interpolation points and to evaluate how much the quality of an inpainted image depends on selection of sparse interpolation points and their sparsity. Compared diffusions are homogeneous diffusion, biharmonic smoothing, triharmonic smoothing, absolute minimal Lipschitz extension, total variation diffusion, isotropic nonlinear diffusion, fourth order isotropic nonlinear diffusion and anisotropic nonlinear diffusion.

Organisation of the paper. In Section 2 a unified model for interpolation is derived. Section 3 specific smoothing operators are described. Experiments and more detailed comparison both quantitatively and qualitatively on PDE-based interpolation

for various sparse images are presented in Section 4. Implications of obtained results are finally discussed in Section 5.

2. A General Model for PDE-based Interpolation

The goal behind PDE-based interpolation is to compute reconstruction of the original image from the pixels that are known without deteriorating them. The image can be reconstructed from sparse data to a relative accuracy using PDE-based interpolation.

The notion of diffusion for most is known from physical context. It is a process that equilibrates concentration differences without creating or destroying mass. This idea carries over to image processing tasks and we will formulate it in a continuous setting.

Let $\Omega \subseteq \mathbb{R}^n$ denote an *n*-dimensional image domain. We want to recover some unknown scalar-valued function $v : \Omega \to \mathbb{R}$, from which we know its values on some subset $\Omega_1 \subseteq \Omega$. Our goal is to find an interpolating function $u : \Omega \to \mathbb{R}$ that is smooth and close to v in Ω/Ω_1 and identical to v in Ω_1.

We may embed this problem in an evolution setting with some evolution parameter (the "time") $t \geq 0$. Its solution $u(x,t)$ gives the desired interpolating function as its steady state $(t \to \infty)$. We initialise the evolution with some function $f : \Omega \to \mathbb{R}$ that is identical to v on Ω_1 and that is set to some arbitrary value (e.g. to 0) on Ω/Ω_1:

$$f(x) := \begin{cases} v(x) & \text{if } x \in \Omega_1 \\ 0 & \text{else.} \end{cases} \tag{1}$$

We consider the evolution

$$\partial_t u = (1 - c(x))Lu - c(x)(u - f) \tag{2}$$

with *f* as initial value,

$$u(x,0) = f(x), \tag{3}$$

and reflecting (homogeneous Neumann) boundary conditions on the image boundary $\partial\Omega$. The function $c : \Omega \to \mathbb{R}$ is the characteristic function on Ω_1, i.e.

$$c(x) := \begin{cases} 1 & \text{if } x \in \Omega_1 \\ 0 & \text{else,} \end{cases} \tag{4}$$

and L is some elliptic differential operator. The idea is to solve the steady state equation

$$\big(1 - c(x)\big) Lu - c(x)\big(u - f\big) = 0 \tag{5}$$

with reflecting boundary conditions. In Ω_1 we have $c(x) = 1$ such that the interpolation condition $u(x) = f(x) = v(x)$ is fulfilled. In Ω/Ω_1 it follows from $c(x) = 0$ that the solution has to satisfy $Lu = 0$. This elliptic PDE can be regarded as the steady state of the evolution equation

$$\partial_t u = Lu \tag{6}$$

with Dirichlet boundary conditions given by the interpolation data on Ω_1.

3. Specific Smoothing Operators

Regarding the elliptic differential operator L, many possibilities exist. The simplest and the best investigated one uses the Laplacian $Lu := \Delta u$ leading to *linear isotropic diffusion* [32]:

$$\partial_t u = \Delta u, \tag{7}$$

also called *homogeneous diffusion*. Linear diffusion is well-posed and can be very well axiomatically justified, it fulfils a maximum-minimum principle and requires no additional parameters (besides t). Diffusion is also its main problem, it is not only good in removing noise and small scale details but it also removes important image features like edges and corners.

A prototype for a higher order differential operator is the biharmonic operator $Lu := -\Delta^2 u$ giving the biharmonic smoothing evolution

$$\partial_t u = -\Delta^2 u. \tag{8}$$

Using it for interpolation comes down to thin plate spline interpolation [33], a rotationally invariant multidimensional generalisations of cubic spline interpolation. Note that only the second-order differential operators allow a maximum--minimum principle where the values of u stay within the range of the values of f in Ω_1.

The multidimensional generalisation of quintic spline interpolation leads to *triharmonic smoothing* based on $Lu := \Delta^3 u$:

$$\partial_t u = \Delta^3 u. \tag{9}$$

Results of triharmonic smoothing suffer from over- and undershoots because of violation of an extremum principle.

A second order PDE that uses the second order directional derivative $Lu := \partial_{\eta\eta} u$ in gradient direction $\eta := \nabla u / |\nabla u|$ as proposed in [10] is *absolute minimal Lipschitz extension (AMLE)* model:

$$\partial_t u = \partial_{\eta\eta} u. \tag{10}$$

It has been axiomatically justified for interpolation purposes. This differential operator acts anisotropically since it interpolates only in the direction of the gradient.

Total variation diffusion (TV) [34], [35] that uses $Lu := div\left(\nabla u / |\nabla u|\right)$ is a nonlinear diffusion filter that is based on the equation

$$\partial_t u = div\left(\frac{\nabla u}{|\nabla u|}\right) \tag{11}$$

and uses the unbounded diffusivity

$$g\left(|\nabla u|^2\right) = \frac{1}{|\nabla u|}. \tag{12}$$

It is well-posed since it preserves the shape of some objects and it leads to constant signals in finite time. Main disadvantage is that unbounded diffusivity (12) creates numerical problems and its approximation may not reproduce some of the nice theoretical properties.

While diffusion has effect of smoothing it also destroys sharp edges which we want to keep. In order to suppress diffusion at edges researchers developed nonlinear diffusion. *Nonlinear isotropic diffusion* goal is to reduce smoothing at edges where $|\nabla u|$ is large by using operator $Lu := div\left(g\left(|\nabla u|^2\right)\nabla u\right)$. This gives [36]:

$$\partial_t u = div\left(g\left(|\nabla u|^2\right)\nabla u\right) \tag{13}$$

where the diffusivity function *g* is decreasing in its argument, since the one may e.g. choose the *Charbonnier diffusivity* [37]

$$g\left(\left|\nabla u\right|^2\right) = \frac{1}{\sqrt{1+\frac{\left|\nabla u\right|^2}{\lambda^2}}} \tag{14}$$

with some contrast parameter $\lambda > 0$. Since (13) uses a scalar-valued diffusivity we name this process *isotropic*. Since g depends only on the magnitude of the gradient of u, the diffusion is suppressed equally in all directions. Ideally it would be smoothing along edges, while suppressing it across edges. Problem of the nonlinear isotropic diffusion is that at the edge noisy edge will remain noisy since the diffusivity is reduced.

Higher order nonlinear diffusion considered in this paper is related to differential operator based on Laplacian $Lu := -\Delta\left(g\left(\left|\Delta u\right|^2\right)\Delta u\right)$ proposed by [37], [38]. *Fourth order nonlinear diffusion equation* is

$$\partial_t u = -\Delta\left(g\left(\left|\Delta u\right|^2\right)\Delta u\right) \tag{15}$$

where diffusivity function g is

$$g\left(\left|\Delta u\right|^2\right) = \frac{1}{\sqrt{1+\frac{\left|\Delta u\right|^2}{\lambda^2}}} \tag{16}$$

since the highest derivative operator appearing in (15) is four, it will be denoted as *fourth order Charbonnier diffusion*. The Laplacian of an image at a pixel is zero if the image is planar in its neighbourhood. The fourth-order PDE model uses a piecewise planar image to interpolate. It can reduce noise and preserve edges but it is influenced by speckle effect.

Real anisotropic behaviour is possible when the diffusion coefficient is replaced with a diffusion tensor. As a prototype for nonlinear anisotropic diffusion filtering we consider *edge-enhancing diffusion (EED)* [39]. The idea is to reduce smoothing across edges while still permitting diffusion along them. The EED diffusion tensor has one eigenvector parallel to ∇u_σ. The associated eigenvalue is given by $g\left(\left|\nabla u_\sigma\right|^2\right)$ with a diffusivity function such as (14). The other eigenvectors are orthogonal to ∇u_σ and have corresponding eigenvalues equal to 1. If we use the convention to extend a scalar-valued function $g(x)$ to a matrix-valued function $g(A)$ by applying g to the eigenvalues on A and leaving the eigenvectors unchanged, then EED can be formally linked to $Lu := div\left(g\left(\nabla u_\sigma \nabla u_\sigma^{\mathrm{T}}\right)\nabla u_\sigma\right)$. Hence, its evolution is governed by

$$\partial_t u = div\left(g\left(\nabla u_\sigma \nabla u_\sigma^{\mathrm{T}}\right)\nabla u_\sigma\right). \tag{17}$$

Anisotropic diffusion has higher computational cost but anisotropy allows EED to remove noise from noisy edges while interpolating.

4. Experiments on Interpolation

Experiments illustrate the use of the different smoothing operators for scattered data interpolation. In order to evaluate the potential of the preceding PDEs for scattered data interpolation, they have been discretised with central finite differences in space. For the diffusion equations, a semi-implicit time discretisation with SOR as solver for the linear systems of equations is performed, while for AMLE, second order nonlinear diffusion, biharmonic and triharmonic smoothing an explicit scheme is used. For TV discretisation forward-backward proximal splitting algorithm is used. Nonlinear PDEs can be solved using explicit or semi-implicit schemes that are based on finite-difference approximations of the original formulation [39]. Runtimes for a non-optimised C implementation on a Quad-Core Xeon 2.8 GHz and 12 GB of DDR3 memory desktop range between a few seconds and several minutes for a 256×256 image. All test were run on zoom into the test image *lena*, 256×256 pixels, Figure 1.

To evaluate results, several objective error measures between the interpolated image $u_{i,j}$ and the original image $v_{i,j}$ have been computed.

Fig. 1 – Zoom into the test image *lena*, 256×256 pixels.

Average absolute error (AAE):

$$AAE(u,v) := \frac{1}{N}\sum_{i,j}\left|u_{i,j} - v_{i,j}\right| \tag{18}$$

where N denotes the number of pixels. Smaller AAE means smaller difference of decoded and original image.

Mean squared error (MSE):

$$MSE(u,v) := \frac{1}{N}\sum_{i,j}\left|u_{i,j} - v_{i,j}\right|^2 \tag{19}$$

MSE is a measure of noise power. Square of the error attenuates small differences between two pictures but increases big differences. Smaller values of MSE mean smaller error, i.e. smaller differences of decoded and original image.

Signal to noise ratio (SNR):

$$SNR(u,v) := 10 \cdot \log_{10}\left(\frac{\sigma^2(g)}{MSE}\right) \tag{20}$$

SNR is a ratio of the average signal strength and average strength of present noise, i.e. in this case differences from the original picture. Larger value of SNR means bigger similarity with the original image.

The amplitude of picture elements has a range MAX, where q is the number of bits needed to display the amplitude of the original image elements. MSE does not take into account the MAX, therefore we introduce *peak signal to noise ratio (PSNR)*:

$$PSNR(u,v) := 10\log_{10}\frac{MAX}{MSE} = 20\log_{10}\frac{MAX}{\sqrt{MSE}}. \tag{21}$$

PSNR is the logarithmic ration between the maximum signal strength and noise strength. Unit of measure is decibel (dB). The value of PSNR = 0 means there is no similarity between the tested pictures, while infinite value of PSNR means that the two pictures are identical.

In these experiments greyscale image is interpolated from grey values of scattered interpolation points that are stored as sparse image which is black everywhere except in the locations of sparse scattered data, a fraction of the image data. Approximately only 2 *percent* of significant pixels are being kept as scattered interpolation points. It is a difficult task how to decide which pixels should be retained in a

sparse set and it is not a topic of this paper. Analytic approach about the optimal data selection it is discussed in [40]. There are various ways to choose the subset of pixels that will be stored as sparse image. Six strategies of removing less significant pixels from the image are considered: random, regular, corners and three adaptive ones. Locations of the preserved pixels are the inpainting mask.

In Figure 2 (a) sparse image is depicted where approximately 2 *percent* of all pixels from test image have been chosen randomly as scattered interpolation points. The density and position of sparse data is better visible in interpolation mask Figure 2 (b). Therefore, interpolation mask will be displayed together with sparse image for all posterior experiments. We observe that homogeneous diffusion, Figure 2 (c), creates singularities at interpolation points. This can be avoided with interpolation using biharmonic smoothing Figure 2 (d). It gives fairly good results, but suffers

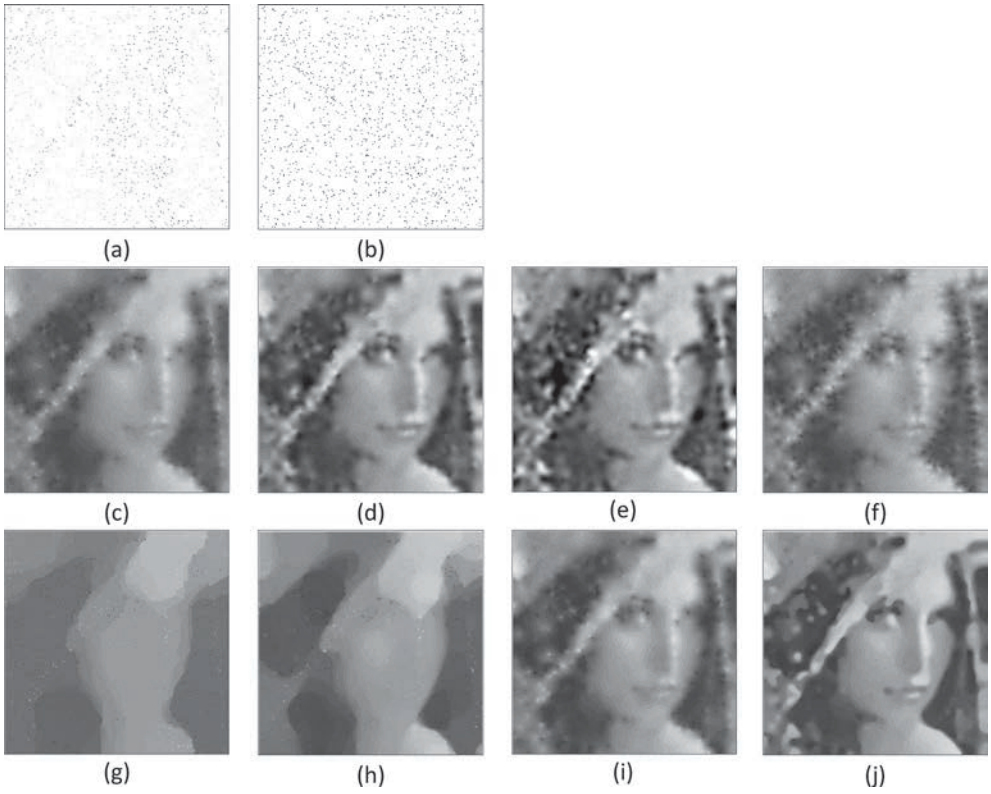


Fig. 2 – (a) Random sparse image (1369 scattered interpolation points). (b) Random mask. Interpolation by: (c) Homogeneous diffusion (t = 1000, 20 iterations). (d) Biharmonic smoothing (t = 0.031, 159540 iterations). (e) Triharmonic smoothing (t = 0.0039, 5657500 iterations). (f) AMLE (t = 0.07, 10000 iterations). (g) TV (t = 0.25, α = 50, 5000 iterations). (h) Charbonnier diffusion (t = 1000, λ = 0.1, 390 iterations). (i) 4th order Charbonnier diffusion (t = 0.031, λ = 1, 345630 iterations) (j) EED (t = 1000, λ = 0.14, σ = 1.1, 1290 iterations).

from over- and undershoots near edges due to the violation of an extremum principle (see e.g. the shoulder). These limitations become even more obvious for triharmonic smoothing Figure 2 (e). In Figure 2 (f) we observe that in spite of a number of favourable theoretical properties, AMLE does not give a good result as it would be expected from its axiomatic justification. The worst result is with TV, Figure 2 (g), reconstructed image appears segmented, it has reduced contrast and suffers from singularities. Nonlinear isotropic diffusion Figure 2 (h) does not give an improvement. While nonlinear isotropic diffusion may allow discontinuities, its interpolant is too flat and tends to keep many interpolation points as isolated singularities. 4th order Charbonnier diffusion Figure 2 (i) gives much better result than Charbonnier diffusion, but it suffers from violation of maximum–minimum principle and it has singularities at interpolation points. EED Figure 2 (j) gives the best result. Its ability to smooth along edges seems to be very beneficial for avoiding singularities at interpolation points. Moreover, this second-order PDE respects a maximum–minimum principle, such that the solution is within the greyscale bounds of the interpolation points.

Table 1 – Average absolute errors (AAE), mean squared error (MSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) of restored images in Figure 2 (c)-(j) from random sparse image Figure 2 (a). The best result in each column is highlighted.

PDE method	AAE	MSE	SNR	PSNR
homogeneous diffusion (7)	16.47	562.98	4.81	20.63
biharmonic smoothing (8)	14.25	517.58	6.67	20.99
triharmonic smoothing (9)	17.24	754.23	5.71	19.36
AMLE (10)	17.06	597.86	4.60	20.37
TV (11)	29.07	1372.13	–2.23	16.76
Charbonnier diffusion (13)	21.18	983.72	2.34	18.41
4th order Charbonnier diffusion (15)	14.80	516.48	5.83	21.00
EED (17)	13.12	483.41	6.88	21.29

The quantitative error analysis for Figure 2 is presented in Table 1. Higher–order interpolation operators that violate maximum–minimum principle may create over- and undershoots. Therefore, at locations where they lead to grey values outside the interval [0, 255], they have been truncated.

Second sparsification strategy is a simplistic method because pixels in a test image are chosen uniformly as scattered interpolation points on a sparse regular grid, Figure 3 (a). Almost all tested PDSs benefit from uniformly dense scattered interpola-

tion points. EED provide the best result, Figure 3 (j), followed by biharmonic smoothing, Figure 3 (f), and triharmonic smoothing, Figure 3 (e). The most significant improvement is visible for triharmonic smoothing despite the over- and undershoots (see Lena's shoulder, cheek). 4th order Charbonnier diffusion smooths nicely but suffers from strong singularities at edges (see Lena's cheek, nose, hat) and low contrast. The quality diminishes with increased number of singularities, first homogeneous diffusion Figure 3(c)

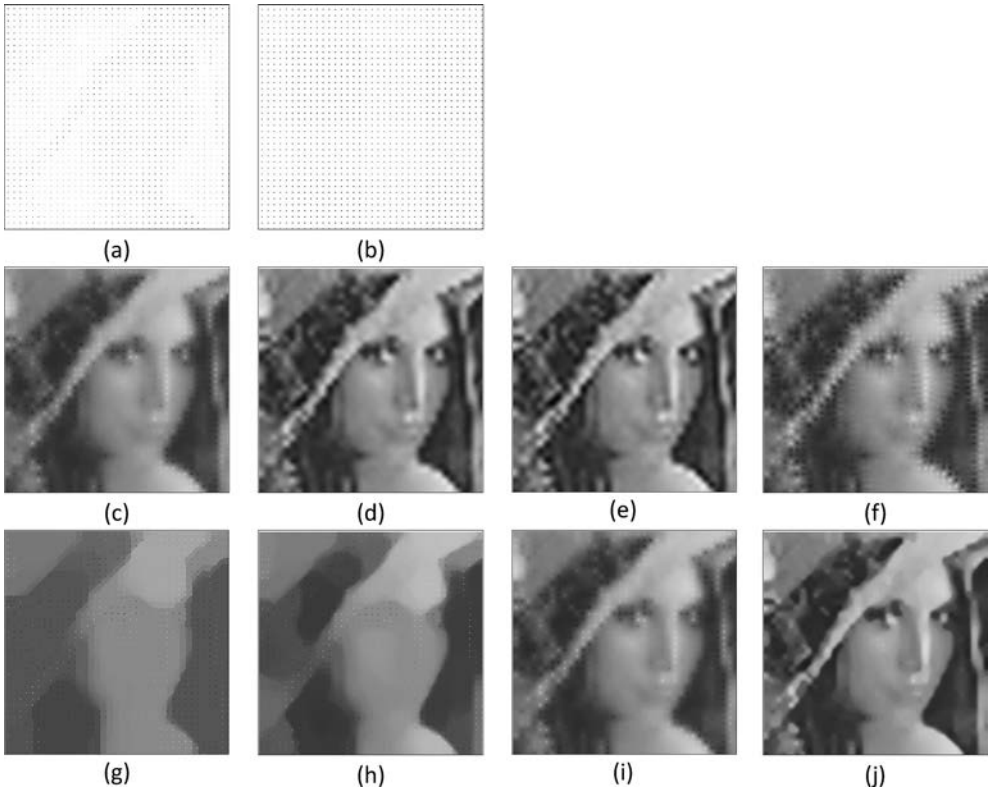


Fig. 3 – (a) Regular grid sparse image (1369 scattered interpolation points). (b) Regular grid mask. Interpolation by: (c) Homogeneous diffusion (t = 1000, 20 iterations). (d) Biharmonic smoothing (t = 0.031, 4890 iterations). (e) Triharmonic smoothing (t = 0.0039, 90450 iterations). (f) AMLE (t = 0.021, 5000 iterations). (g) TV (t = 0.25, α=50, 5000 iterations). (h) Charbonnier diffusion (t = 1000, λ = 0.1, 350 iterations). (i) 4th order Charbonnier diffusion (t = 0.031, λ = 0.1, 756030 iterations). (j) EED (t = 1000, λ = 0.17, σ = 0.5, 20 iterations).

then AMLE Figure 3 (f). Again TV, Figure 3 (g), and Charbonnier diffusion, Figure 3 (h), give inferior interpolation.

The quantitative error analysis for Figure 3 presented in Table 2 coincides with the subjective quality assessment.

Table 2 – Average absolute errors (AAE), mean squared error (MSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) of restored images in Figure 3 (c)-(j) from regular grid sparse image Figure 3 (a). The best result in each column is highlighted.

PDE method	AAE	MSE	SNR	PSNR
homogeneous diffusion (7)	14.56	456.40	6.01	21.54
biharmonic smoothing (8)	12.27	390.21	7.92	22.22
triharmonic smoothing (9)	12.82	417.23	7.79	21.93
AMLE (10)	15.12	479.01	5.78	21.33
TV (11)	28.18	1295.33	–1.95	17.01
Charbonnier diffusion (13)	20.01	841.31	2.74	18.88
4th order Charbonnier diffusion (15)	14.31	449.79	6.18	21.60
EED (17)	11.54	360.40	8.16	22.56

The third strategy for selecting scattered interpolation points is at image locations where the edge changes its direction, namely corners. Corners are relevant for understanding the image content. Corners are sparser features than edges and since the goal is only 2*percent* it seems more reasonable to detect corners because edges would be too extensive in this context. Corners are easy to obtain from corner detection algorithm that uses structure tensor proposed by [41]. The structure tensor gives a gradient-like descriptor of local image structure. The proposed strategy does not require computation of eigenvalues explicitly but the trace and the determinant. Figure 4 (a) shows sparse image where 2 percent of all pixels have been chosen according to the semantically motivated feature selection of the image data by corner detection and threshold in order to get desired number of scattered interpolation points. There is a drawback; corners are a very sparse feature and all tested PDEs cannot cope with that and derive unsatisfactory result. It is interesting to observe that biharmonic smoothing, Figures 4 (d), triharmonic smoothing, Figures 4 (e) and Charbonnier diffusion fourth order, Figures 4 (i) fail to interpolate complete image but in the area where interpolation points are dense (see Lena's eyes) produce some significant results. The rest of PDEs fills in the empty region as flat surface but the result is not convincing. This experiment shows that some of tested PDEs are sensitive to density of scattered interpolation points and others that will interpolate no matter how sparse data is.

The quantitative error analysis for Figure 4 is presented in Table 3. It is interesting to notice that only in this experiment quantitative analysis shows significant discrepancy between error measures. Biharmonic smoothing is singled out as the best by SNR contrary to AAE, MSE and PSNR that choose EED over all compared PDEs and classify biharmonic smoothing among the inferior ones. It is also first and only experiment where Charbonnier diffusion and TV are not the worst but it is only due to its ability to fill in the empty region as flat surface with average grey value.

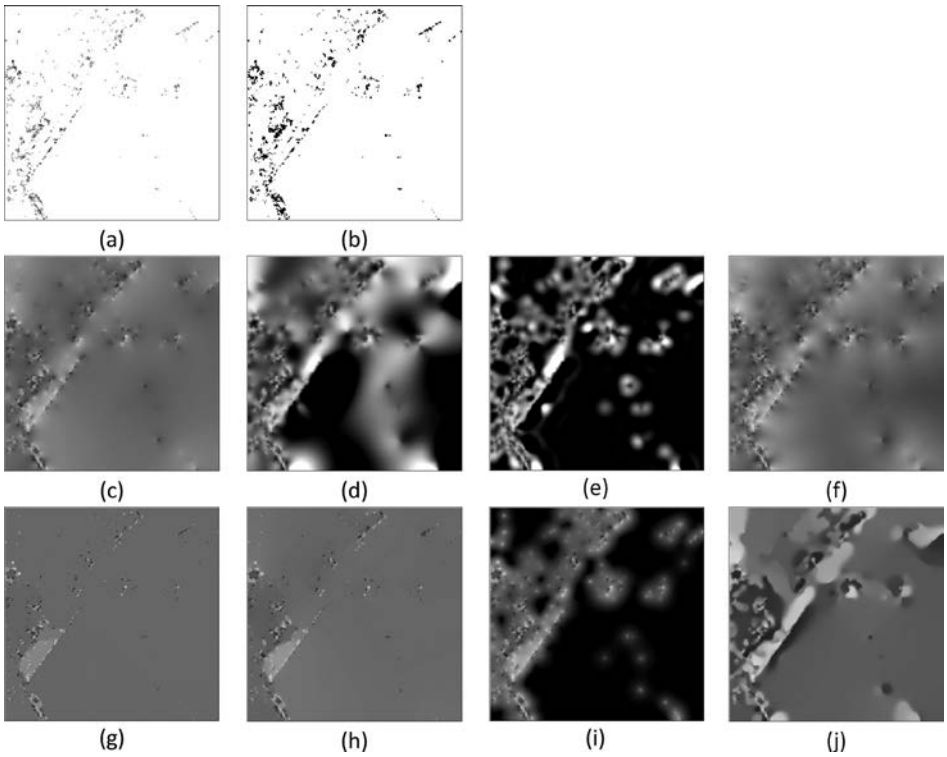


Fig. 4 – (a) Corner sparse image (1422 scattered interpolation points). (b) Corner mask. In- terpolation by: (c) Homogeneous diffusion (t = 1000, 40 iterations). (d) Biharmonic smoothing (t = 0.031, 1500000 iterations). (e) Triharmonic smoothing (t = 0.0039, 500000 iterations). (f) AMLE (t = 0.024, 10000 iterations). (g) TV (t = 0.25, α = 50, 5000 iterations). (h) Charbonnier diffusion (t = 1000, λ = 0.3, 230 iterations). (i) 4th order Charbonnier diffusion (t = 0.031, λ = 0.1, 1000000 iterations). (j) EED (t = 1000, λ = 0.09, σ = 1, 5000 iterations).

Table 3 – Average absolute errors (AAE), mean squared error (MSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) of restored images in Figure 4 (c)-(j) from corner sparse image Figure 4 (a). The best result in each column is highlighted.

PDE method	AAE	MSE	SNR	PSNR
homogeneous diffusion (7)	42.57	2477.13	–9.32	14.19
biharmonic smoothing (8)	54.30	4877.85	0.33	11.25
triharmonic smoothing (9)	84.62	10625.29	–1.84	7.87
AMLE (10)	42.49	2517.62	–6.72	14.12
TV (11)	44.17	2620.04	–16.12	13.95
Charbonnier diffusion (13)	43.50	2568.43	–14.31	14.03
4th order Charbonnier diffusion (15)	84.87	10650.39	–4.02	7.86
EED (17)	39.18	2363.10	–3.94	14.40

First adaptive sparsification of the image data is by means of the triangulation from B-tree triangular coding (BTTC) [29]. BTTC is a simple and efficient algorithm for adaptive pixel selection.

It works by dividing the image's surface into right triangles. Whenever a triangle's linear interpolation is too far from the original image by some threshold, the triangle is subdivided into two right triangles. The pixels selected are then close to the areas of high curvature. The BTTC method has efficient encoding in a tree structure. Sparse image is depicted in Figure 5 (a) where 2 percent of all pixels have been chosen according to the adaptive sparsification of the image data by BTTC. It is not surprising that EED, Figure 5 (j), gives the best result, followed by biharmonic smoothing, Figure 5 (d), and homogeneous diffusion, Figure 5 (c). 4th order Charbonnier diffusion, Figure 5 (i), gives better result than AMLE, Figure 5 (f),

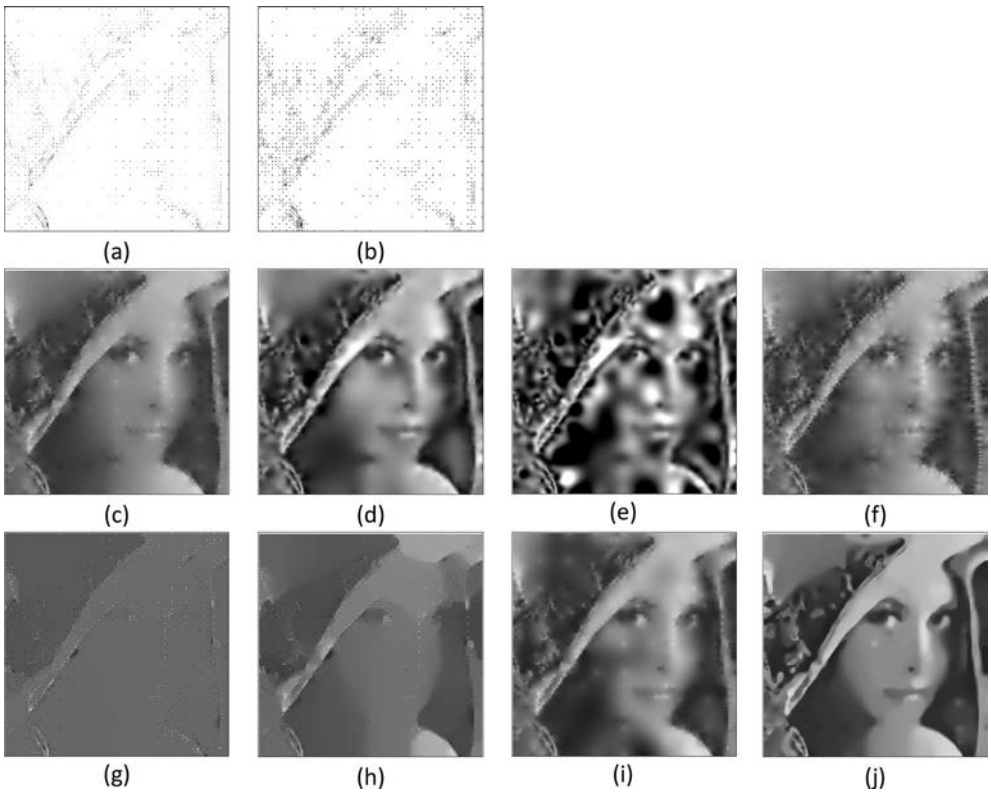


Fig. 5 – (a) BTTC sparse image (1365 scattered interpolation points). (b) BTTC adaptive mask. Interpolation by: (c) Homogeneous diffusion (t = 1000, 20 iterations). (d) Biharmonic smoothing (t = 0.031, 738410 iterations). (e) Triharmonic smoothing (t = 0.0039, 37014650 iterations). (f) AMLE (t = 0.1, 10000 iterations). (g) TV (t = 0.25, α = 50, 5000 iterations). (h) Charbonnier diffusion (t = 1000, λ = 0.1, 510 iterations). (i) 4th order Charbonnier diffusion (t = 0.031, λ = 1, 1000000 iterations). (j) EED (t = 1000, λ = 0.1, σ = 1, 20 iterations).

because the image looks smoother and AMLE has very strong singularities. Triharmonic smoothing, Figure 5 (e), didnot interpolate whole image. A quantitative error analysis for Figure 5 is presented in Table 4. Although, EED gives the best result AAE is worse than for random sparse image but MSE, SNR and PSNR confirm betterment perceived by subjective qualitative assessment. 4th order Charbonnier diffusion is assessed better than AMLE by visual examination that concurs only with SNR, but differs from other quantitative error analysis measures.

Table 4 – Average absolute errors (AAE), mean squared error (MSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) of restored images in Figure 5 (c)-(j) from BTTC sparse image Figure 5 (a). The best result in each column is highlighted.

PDE method	AAE	MSE	SNR	PSNR
homogeneous diffusion (7)	17.29	450.39	5.33	21.59
biharmonic smoothing (8)	14.90	400.42	8.55	22.11
triharmonic smoothing (9)	35.95	2610.51	2.28	13.96
AMLE (10)	17.42	476.56	5.25	21.35
TV (11)	42.67	2421.12	–13.06	14.29
Charbonnier diffusion (13)	31.11	1331.63	–3.42	16.89
4th order Charbonnier diffusion (15)	19.05	665.27	5.41	19.90
EED (17)	12.18	296.75	9.19	23.41

In EEDC adaptive triangulation from [21] is used, linear interpolation within a triangle is replaced by EED on whole image together with different amendments. In sparse image Figure 6 (a) where 2percent of all pixels have been chosen according to the adaptive sparsification of the image data by EEDC. As expected the quality of interpolation by EED, Figure 6 (j), improved significantly when EED was included in adaptive sparsification process and is appreciable better than rest of the interpolations. Subsidiary result achieves biharmonic smoothing, Figure 6 (d), coming immediately in quality is 4th order Charbonnier diffusion, Figure 6 (i), then interpolation by homogeneous diffusion, Figure 6 (c), and AMLE, Figure 6 (f).

A quantitative error analysis for Figure 6 is presented in Table 5 and conforms to subjective quality assessment.

The quality of EED based inpainting significantly improved when sparification was adapted to the EED. Therefore it is interesting to see if and how much the quality

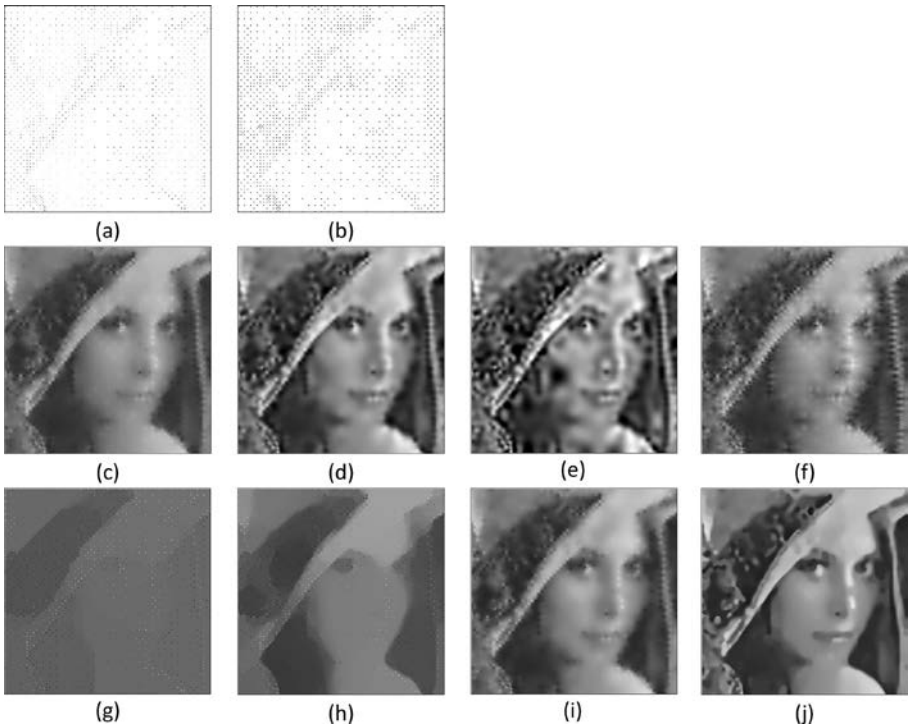


Fig. 6 – (a) EEDC sparse image (1362 scattered interpolation points). (b) EEDC adaptive mask. Interpolation by: (c) Homogeneous diffusion (t = 1000, 20 iterations). (d) Biharmonic smoothing (t = 0.031, 165750 iterations). (e) Triharmonic smoothing (t = 0.0039, 5000000 iterations). (f) AMLE (t = 0.09, 10000 iterations). (g) TV (t = 0.25, α = 50, 680 iterations). (h) Charbonnier diffusion (t = 1000, λ = 0.1, 330 iterations). (i) 4th order Charbonnier diffusion (t = 0.031, λ = 1, 570620 iterations). (j) EED (t = 1000, λ = 0.18, σ = 1.3, 2000 iterations).

Table 5 – Average absolute errors (AAE), mean squared error (MSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) of restored images in Figure 6 (c)-(j) from EEDC sparse image Figure 6 (a). The best result in each column is highlighted.

PDE method	AAE	MSE	SNR	PSNR
homogeneous diffusion (7)	14.82	372.53	6.57	22.42
biharmonic smoothing (8)	11.77	254.85	10.06	24.07
triharmonic smoothing (9)	21.80	1018.65	4.57	18.05
AMLE (10)	15.20	412.70	6.34	21.99
TV (11)	40.50	2225.50	–12.42	14.657
Charbonnier diffusion (13)	24.73	985.51	–0.30	18.19
4th order Charbonnier diffusion (15)	12.55	309.49	8.06	23.22
EED (17)	9.62	194.90	10.94	25.23

of reconstructed image will benefit when adaptive sparsification is adapted to some other diffusion and might even outperform EED. Immediate choice is homogeneous diffusion that needs only 20 iterations and it has relatively good reconstruction quality compared to other PDEs therefore is the perfect candidate for this experiment. Modified EEDC adaptive sparsification is used in sixth strategy where EED is substituted with homogeneous diffusion. This adaptive sparsification strategy is denoted as HDC. As expected best interpolation is obtained with EED, Figure 7 (j), but homogeneous diffusion, Figure 7 (c), is not the second best. It is not the biharmonic smoothing, Figure 7 (d), that outperforms homogeneous diffusion, surprisingly it is 4th order Charbonnier diffusion, Figure 7 (i) that gives convincing result. However unsatisfactory, image reconstructed from HDC sparse image gives the best result out of all interpolations by Charbonnier diffusion and is shown in Figure 7 (h).

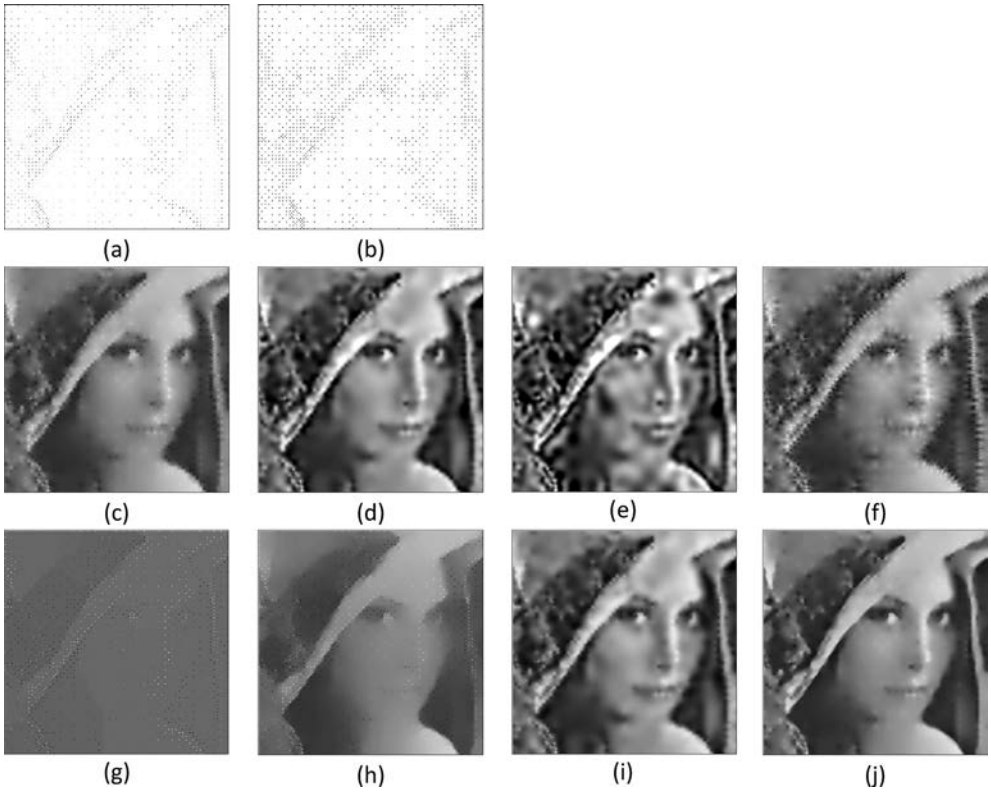


Figure 7: (a) HDC sparse image (1365 scattered interpolation points). (b) HDC adaptive mask. Interpolation by: (c) Homogeneous diffusion (t = 1000, 20 iterations). (d) Biharmonic smoothing (t = 0.031, 172310 iterations). (e) Triharmonic smoothing (t = 0.0039, 4501450 iterations). (f) AMLE (t = 0.07, 10000 iterations). (g) TV (t = 0.25, α = 50, 5000 iterations). (h) Charbonnier diffusion (t = 1000, λ = 1, 130 iterations). (i) 4th order Charbonnier diffusion (t = 0.031, λ = 1, 635740 iterations). (j) EED (t = 1000, λ = 2.5, σ = 1.7, 50 iterations).

Corresponding quantitative error analysis for Figure 7 is presented in Table 6.

Table 6 – Average absolute errors (AAE), mean squared error (MSE), signal to noise ratio (SNR) and peak signal to noise ratio (PSNR) of restored images in Figure 7 (c)-(j) from HDC sparse image Figure 7 (a). The best result in each column is highlighted.

PDE method	AAE	MSE	SNR	PSNR
homogeneous diffusion (7)	12.47	269.74	8.84	23.82
biharmonic smoothing (8)	14.84	396.35	9.21	22.15
triharmonic smoothing (9)	24.59	1210.61	4.93	17.30
AMLE (10)	13.74	347.99	7.93	22.72
TV (11)	42.39	2387.91	–13.69	14.35
Charbonnier diffusion (13)	22.28	724.75	2.43	19.52
4th order Charbonnier diffusion (15)	11.59	247.34	9.83	24.20
EED (17)	10.77	216.23	10.77	24.78

In all experiments the average absolute error gives a ranking that corresponds well with visual impression. Plot depicted in Figure 8 shows the value of AAE for certain combinations of sparse images and PDEs. In all quantitative and qualitative

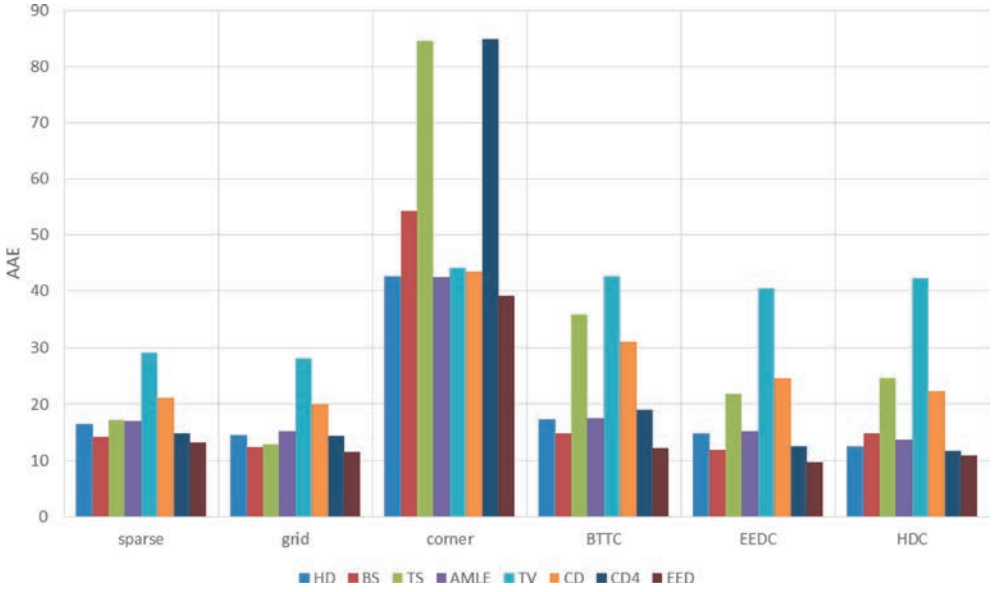


Fig. 8 – Comparison of tested smoothing operators over a choice of sparse masks for the test image *lena* evaluated with AAE.

assessments EED gives the best scattered data interpolation. These findings are also in accordance with results from [13], [26] where EED proved to be the PDE of choice for interpolation. The quality of reconstruction from inpainting with homogeneous diffusion, biharmonic smoothing and 4th order Charbonnier diffusion vary considerably. The higher order operators (biharmonic, triharmonic smoothing and 4th order Charbonnier diffusion) show visible fluctuations due to over- and undershoots at edges. This is characteristic for higher-order differential operators that violate a maximum-minimum principle. The second- order operators (linear diffusion, AMLE, TV, Charbonnier interpolation) suffer from the fact that the information in the selected pixels seems to be too sparse to create regular edge contours that do not appear blurry. We observe that TV is not good for scattered data interpolation although it works quite well for denoising.

5. Conclusions

The paper addresses the reconstruction problem of an image from their scattered interpolation points and recovery of missing data. PDE-based inpainting ideas have been driven to the extreme by storing only a sparse set of all pixels and interpolating the missing data by various PDEs. The choice of appropriate inpainting operators has been reviewed. Six sparse images were inpainted with different PDEs and results were compared both qualitatively and quantitatively.

Experiments indicate that despite of its simplicity linear diffusion can give good interpolation results if the interpolation data is chosen carefully. Linear diffusion is capable of inpainting sparse images Nevertheless, for compression purposes, but singularities at the interpolation points exist and the image is blurred. It does not only blur all structures in equal amount but also dislocates them.

Biharmonic smoothing increases smoothness around interpolation points, but it suffers from over- and undershoots due to a lack of max-min principle. It was shown that biharmonic smoothing can offer a satisfactory reconstruction quality depending on a selection of scattered interpolation points. Triharmonic smoothing did not provide adequate results when used for interpolation of sparse images, where the disadvantages of the violation of a maximum minimum principle become more evident. The best result was obtained with regular grid sparse image where density of interpolation points was minimal.

AMLE and TV do not live up to its expectations. AMLE suffers from the problem that interpolation points are too sparse to create sharp edges and has largest number of singularities at interpolation points near discontinuities out of all interpolation

methods investigated. In case of total-variation diffusion (TV) the behavior of standard edge-preserving regularization techniques is not adapted to sparse interpolation for it produces singular points. Useful properties like its tendency to yield segmentation-like results and linear contrast reduction that gave fairly good results when denoising did not work well for sparse inpainting. It could not restore smoothly varying region and performed consistently bad.

Charbonnier diffusion fails because it hardly respects the boundary conditions, interpolant is too flat, it does not smooth edges and it rounds corners. For all tested sparse images reconstructed image was regularly unsatisfactory. 4th order Charbonnier diffusion is proposed to optimize the trade-off between smoothing and edge preservation. Reconstructed image is piecewise planar, it looks more natural but it tends to reduce contrast and leaves the processed images with isolated black and white singularities at interpolation points which may be characterized as pixels whose intensity values are either much larger or smaller than those of their neighboring pixels.

For scattered data interpolation as particularly useful PDE edge-enhancing anisotropic diffusion (EED) is identified. It uses a diffusion tensor that allows smoothing along discontinuities while inhibiting smoothing across them. Edge-enhancing-anisotropic diffusion enjoys better reconstruction properties and remarkable edge-reconstruction performance. Experiments show the suitability of EED to restore geometrical information from few pixels. It has a higher computational cost but also better results. EED is the best one that can cope with ”bad” interpolation points.

At very high sparsity levels it is noticeable how a scattered interpolation points selection can have a strong influence on the interpolation result. From results with corner mask it can be concluded that the reconstruction quality highly benefits from denser point selection. In case of corner sparsification it is conclusive that not even the robust PDE like EED did not succeed to interpolate poorly chosen sparse interpolation points. The density of pixels should be adapted to the underlying image structure in order to allow the best interpolation quality. Nevertheless, for compression purposes, good interpolation quality is not sufficient if the resulting set of image data is too computationally expensive to encode.

References

[1] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester, ‘Image inpainting’, in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, New Orleans, LI, 2000, pp. 417–424.

[2] F. Bornemann and T. März, ‘Fast image inpainting based on coherence transport’, *J. Math. Imaging Vis.*, vol. 28, no. 3, pp. 259–278, Jul. 2007.

[3] T. F. Chan and J. Shen, 'Non-texture inpainting by curvature-driven diffusions', *J. Vis. Commun. Image Represent.*, vol. 12, no. 4, pp. 436–449, 2001.

[4] H. Grossauer and O. Scherzer, 'Using the complex Ginzburg–Landau equation for digital impainting in 2D and 3D', in *Scale-Space Methods in Computer Vision*, vol. 2695, D. Griffin and M. Lillholm, Eds. Berlin.

[5] R. March, 'Computation of stereo disparity using regularization', *Pattern Recognit. Lett.*, vol. 8, no. 3, pp. 181–187, Oct. 1988.

[6] D. Tschumperlé and R. Deriche, 'Vector-valued image regularization with PDEs: A common framework for different applications', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 506–516, Apr. 2005.

[7] H. A. Aly and E. Dubois, 'Image up-sampling using total-variation regularization with a new observation model', *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1647–1659, Oct. 2005.

[8] S. Battiato, G. Gallo, and F. Stanco, 'Smart interpolation by anisotropic diffusion', in *Proc. Twelfth International Conference on Image Analysis and Processing*, 2003, pp. 572–577.

[9] A. Belahmidi and F. Guichard, 'A partial differential equation approach to image zoom', in *Proc. 2004 IEEE International Conference on Image Processing*, Singapore, 2004, vol. 1, pp. 649–652.

[10] V. Caselles, J. –. Morel, and C. Sbert, 'An axiomatic approach to image interpolation', *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 376–386, Mar. 1998.

[11] F. Malgouyres and F. Guichard, 'Edge direction preserving image zooming: A mathematical and numerical analysis', *SIAM J. Numer. Anal.*, vol. 39, no. 1, pp. 1–37, 2001.

[12] A. Roussos and P. Maragos, 'Vector-valued image interpolation by an anisotropic diffusionprojection PDE', in *Scale Space and Variational Methods in Computer Vision*, vol. 4485, F. M. Sgallari and N. Paragios, Eds. Berlin.

[13] J. Weickert and M. Welk, 'Tensor field interpolation with PDEs', in *Visualization and Processing of Tensor Fields*, J. Weickert and H. Hagen, Eds. Berlin: Springer Berlin Heidelberg, 2006, pp. 315–325.

[14] G. Facciolo, F. Lecumberry, A. Almansa, A. Pardo, V. Caselles, and B. Rougé, 'Constrained anisotropic diffusion and some applications', in *Proc. 2006 British Machine Vision Conference*, Edinburgh, Scotland, 2006, pp. 1049–1058.

[15] A. Solé, V. Caselles, G. Sapiro, and F. Arandiga, 'Morse description and geometric encoding of digital elevation maps', *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1245–1262, Sep. 2004.

[16] Z. Xie, W. R. Franklin, B. Cutler, M. A. Andrade, M. Inanc, and D. M. Tracy, 'Surface compression using over-determined Laplacian approximation', in *Proc. SPIE Advanced Signal Processing Algorithms, Architectures, and Implementations XVII*, vol. 5266, T. Luk, Ed. Bellingham: SPIE Press, 2007, p. 66970F–66970F–12.

[17] M. Lillholm, M. Nielsen, and L. D. Griffin, 'Feature-based image analysis', *Int. J. Comput. Vis.*, vol. 52, no. 2–3, pp. 73–95, 2003.

[18] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel, 'Towards PDE-based image compression', in *Variational, Geometric and Level-Set Methods in Computer Vision*, vol. 3752, O. F. Paragios, T. Chan, and C. Schnörr, Eds. Springer Berlin Heidelberg, 2005, pp. 37–48.

[19] H. Zimmer, 'PDE-based image compression using corner information', Dept. of Computer Science, Saarland University, Saarbrücken, Germany, 2007.

[20] H. Köstler, M. Stürmer, C. Freundl, and U. Rüde, 'PDE based video compression in real time', Lehrstuhl für Informatik 10, Univ. Erlangen–Nürnberg, Germany, Technical Report 7-11, 2007.

[21] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H. –. Seidel, 'Image compression with anisotropic diffusion', in *Journal of Mathematical Imaging and Vision*, vol. 31, Berlin: Springer US, 2008, pp. 255–269.

[22] M. Mainberger and J. Weickert, 'Edge-Based Image Compression with Homogeneous Diffusion', in *Computer Analysis of Images and Patterns, Proc. 13th International Conference CAIP 2009*, vol. 5702, X. Jiang and N. Petkov, Eds. Münster, Germany.

[23] C. Schmaltz, J. Weickert, and A. Bruhn, 'Beating the quality of JPEG 2000 with anisotropic diffusion', in *Pattern Recognition*, vol. 5748, H. J., Notni, G., Süsse Denzler, Ed. Berlin.

[24] M. Mainberger, S. Hoffmann, J. Weickert, C. H. Tang, D. Johannsen, F. Neumann, and B. Doerr, 'Optimising spatial and tonal data for homogeneous diffusion inpainting', in *Scale Space and Variational Methods in Computer Vision*, vol. 6667, B. 1te. H. R. Bruckstein, A. M. Bronstein, and M. M. Bronstein, Eds. Berlin.

[25] L. Hoeltgen, S. Setzer, and J. Weickert, 'An optimal control approach to find sparse data for Laplace interpolation', in *Energy Minimisation Methods in Computer Vision and Pattern Recognition*, vol. 8081, F. K. Heyden, C. Olsson, M. Oskarsson, and X. C. Tai, Eds. Berlin.

[26] C. Schmaltz, P. Peter, M. Mainberger, F. Ebel, J. Weickert, and A. Bruhn, 'Understanding, optimising and extending data compression with anisotropic diffusion', *Int. J. Comput. Vis.*, vol. 108, no. 3, pp. 222–240, 2014.

[27] P. Strobach, 'Quadtree-structured recursive plane decomposition coding of images', *IEEE Trans. Signal Process.*, vol. 39, no. 6, pp. 1380–1397, Jun. 1991.

[28] G. J. Sullivan and R. J. Baker, 'Efficient quadtree coding of images and video', *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 327–331, May 1994.

[29] R. Distasi, M. Nappi, and S. Vitulano, 'Image compression by B-tree triangular coding', *IEEE Trans. Commun.*, vol. 45, no. 9, pp. 1095–1100, Sep. 1997.

[30] L. Demaret, N. Dyn, and A. Iske, 'Image compression by linear splines over adaptive triangulations', *Signal Process.*, vol. 86, no. 7, pp. 1604–1616, 2006.

[31] S. Bougleux, G. Peyré, and L. Cohen, 'Image compression with anisotropic triangulations', in *Proc. Tenth International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 2343–2348.

[32] T. Iijima, 'Basic theory on normalization of pattern (in case of typical one-dimensional pattern)', *Bull. Electrotech. Lab.*, vol. 26, pp. 368–388, 1962.

[33] J. Duchon, 'Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces', *ESAIM Math. Model. Numer. Anal. – Modélisation Mathématique Anal. Numér.*, vol. 10, no. 3, pp. 5–12, 1976.

[34] L. I. Rudin, S. Osher, and E. Fatemi, 'Nonlinear total variation based noise removal algorithms', *Phys. Nonlinear Phenom.*, vol. 60, no. 1–4, pp. 259–268, 1992.

[35] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazóon, 'Minimizing total variation flow', *Differ. Integral Equ.*, vol. 14, no. 3, pp. 321–360, Mar. 2001.

[36] P. Perona and J. Malik, 'Scale space and edge detection using anisotropic diffusion', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

[37] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, 'Deterministic edge-preserving regularization in computed imaging', *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, 1997.

[38] S. Didas, J. Weickert, and B. Burgeth, 'Stability and local feature enhancement of higher order nonlinear diffusion filtering', in *Pattern Recognition. Lecture Notes in Computer Science*, vol. 3663, R. S. W. Kropatsch and A. Hanbury, Eds. Springer Berlin Heidelberg, 2005, pp. 451–458.

[39] J. Weickert, 'Theoretical foundations of anisotropic diffusion in image processing', *Theor. Found. Anisotropic Diffus. Image Process.*, vol. 11, pp. 221–236, 1996.

[40] Z. Belhachmi, D. Bucur, B. Burgeth, and J. Weickert, 'How to choose interpolation data in images', *SIAM J. Appl. Math.*, vol. 70, no. 1, pp. 333–352, 2009.

[41] W. Förstner and E. Gülch, 'A fast operator for detection and precise location of distinct points, corners and centres of circular features', in *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, 1987, pp. 281–305.

Scientific paper

Recycling the Prints Obtained by Varying the Voltage in LEP Technology

Z. Bolanča[1,*], I. Bolanča Mirković[2,*], I. Majnarić[2,*]

[1]Croatian Academy of Engineering

[2]University of Zagreb, Faculty of Graphic Arts

With the development of digital printing technology, the amount of digital prints in waste paper is rapidly increasing. At the same time, it has been difficult to detach the ElectroInk in liquid electrophotography printing (LEP), from fibers with the same deinking processes used for conventional inks. Because of the different characteristics of ElectroInk, it has been difficult to collect, recycle and deink. The prints obtained by varying the voltage of the processing drum, reverse roller and a squeegee roller in the LEP technology on the mechanism and efficiency of the deinking flotation process and characteristics of the recycled fibers is presented in these article.

Research results indicate that by increasing negative voltage of the processing drum and reverse roller an increase occurs in the dirt count on handsheets obtained from fibres after disintegration, while the results related to the squeegee roller are inversely proportional. An increase is observed, specific for each of the stages, in the area of dirt > 0.04 mm^2 with the increase of negative voltage. The range of dirt sizes that can be removed by flotation is restricted to between 0.050 and 0.150 mm^2. Dirt below 0.04 mm^2 cannot individually resolved, they still adsorb light and hence contribute a greyness or overall loss of brightness to the handsheet. The paper discusses the inter-dependability of the distribution the size and area of dirt on handsheets obtained from the deinking flotation process with regard to the change in voltage of the processing drum, reverse roller and squeegee roller in printing. The ERIC and handsheet brightness results explain the efficiency of the recycling process well and confirm the results obtained by implementing other methods.

Key words:

LEP technology, recycling, image analysis, ERIC, brightness

[*]zbolanca@hatz.hr, ibolanca@grf.hr, i.majnaric@grf.hr

1. Introduction

Digital printing contributes to sustainable development (Gmelin, Senring, 2014). This printing technique, in comparison to some conventional printing techniques has the advantage of having a less harmful influence on the environment mainly due to its ability to be ready faster along with the absence of plate making and its related chemicals, materials, emissions and wastes (Carver, Guidry, 2011; Kadam et al., 2009; Lasage, Schoonenberg, 2010).

With the development of digital printing technology, the amount of digital prints in waste paper is rapidly increasing. At the same time, it has been difficult to detach the ElectroInk in liquid electrophotography printing (LEP), from fibers with the same deinking processes used for conventional inks (Bolanča, 2000; Bolanča et al., 2000; Carre, Mangin 2002; Smitz, Fischer 2004; Fischer, 2008) Because of the different characteristics of ElectroInk, it has been difficult to collect, recycle and deink.

Digital printing equipment manufacturers and the International Association of the Deinking Industry (INGEDE) have different opinions on the deinkability of digital prints. During DRUPA 2008, HP claimed that the 4.0 version of ElectroInk is as deinkable as dry toners. INGEDE issued a press release stating that the claims were incorrect (INGEDE, 2008). Digital printing equipment manufacturers also question the relevance of the single-step laboratory test of the INGEDE 11 method, claiming that it does not reflect the industrial conditions of two-loop deinking (INGEDE 2001; Wripp2010). A revision of the INGEDE 11 method was done with the criteria of the pH after pulping and before the flotation was added (INGEDE, 2012).

The results of the research conducted showing the effect of kneading and the 2nd flotation on significantly reducing the total dirt area to a level below 200 ppm as the proposed laboratory target were published up until DRUPA 2012 (Hewlett-Packard Indigo, 2012)**.** The new generation of ElectroInk produces smaller specks after pulping. Specks which are effectively broken up by dispersion are floatable in a second flotation loop (Hewlett-Packard Indigo, 2012; Putz et al., 2008). Hewlett-Packard is engaged in research to ensure that the deinkability of current recycled paper streams is maintained as the percentage of digital prints in the deinking mill furnish increases (Ng et al., 2009).

The aim of the research is to understand how ink and paper interaction for liquid electrophotography affect the flotation deinking process. HP Indigo plans to test ElectroInk 4.0 under NewPage Duluth recycled paper mill simulated conditions. The composition of furnish for the deinking trial was 5% ElectroInk and 95% mixed office waste recovered paper (Macias, 2010). The pilot scale simulation

showed that a dirt count is similar to that in a mixed office waste furnish. In this case two flotations were conducted.

Both Mittelstadt and co-authors developed modified alkaline deinking chemistry, HPMA (Mittelstadt et al., 2010). Modified chemistry uses traditional chemicals including sodium hydroxide and sodium silicate with the substitution of a non-ionic surfactant, an emulsifier for fatty acid. The result of the chemical composition is to achieve ink particulates with the desired size range and provide the interfacial interaction with foam for removal during flotation.

Ng and a group of co-authors adapted HPES from a lab scale to a pilot scale and achieved excellent recyclability (Ng et al., 2010). During further research Ng and the fellow associates adapted HPES from a lab scale to a pilot scale (Ng et al., 2010 a). The optimization of HPES is assisted by advanced molecular modelling and dynamics simulation. Using HP Indigo print in mixed-office waste has achieved good recyclability on a pilot scale.

Bhattacharyya and several co-authors used near-neutral deinking chemistry – HPES can be extended to obtain good deinking results for traditional offset inks. They found a direct correlation between the chemical effects on the ink-particle speck contamination and its particle size distributions (Bhattacharyya et al., 2011). Below they presented a comparative study of the alkaline and neutral deinking of indirect electrophotography prints with a variety of substrates (Bhattacharyya et al., 2012). The results show that the preferred option of neutral chemistry is always successful in deinking and that alkaline chemistry is complicated.

Zhang and the collaborating authors describe the design of a laboratory scale two-loop deinking process with the inclusion of a low-speed high shear kneading step in between two flotation steps (Zhang et al., 2013). The samples for recycling consisted of mixed office waste and mixtures of mixed office waste with various amounts of liquid electrophotography prints. It was observed that the two–loop process successfully brings the dirt area for mixed office waste with liquid electrophotography prints to a satisfactory level.

The aim of our extensive research was to determine the influence of the conditions in the printing process of indirect electrophotography on the effectiveness of prints recycling (Bolanča Mirković, Bolanča, 2005, Bolanča Mirković et al., 2014). Further research concentrated on the impact of liquid toner density (D 1.20, 1.40, 1.60, 1.80, and 2.00) in combination with different printing substrates on the characteristic of recycled fibres (Bolanča Mirković, et al., 2009). Furthermore, the mechanism of deinking flotation was presented in relation to the changes in temperature of the intermediate cylinder (Bolanča Mirković et al., 2012). By increasing the temperature in the domain from 125 °C, 130 °C, 135 °C, 140 °C and 145 °C, an increase

of dirt area is observed, which in turn influences their elimination in the flotation process. On the other hand, by increasing the temperature, the increase of colour gamut volume of prints is observed, i.e. higher reproduction quality.

The prints obtained by varying the voltage of the processing drum, reverse roller and squeegee roller in the LEP technology on the mechanism and efficiency of the deinking flotation process and characteristics of the recycled fibers is presented in these article.

The results is contributed to the explanation of the printing process conditions on the deinkability of prints and can be applied in the area of reused materials accepting the cradle-to-cradle concept which takes into consideration the postulates of ecological sustainable development.

2. Materials and methods

The samples were made on the Turbo Stream HP Indigo based LEP technology. HP Indigo printing is an offset process which involves the transfer of a liquid toner, from the imaging drum to a heated offsetting blanket and thence to the printing substrate.

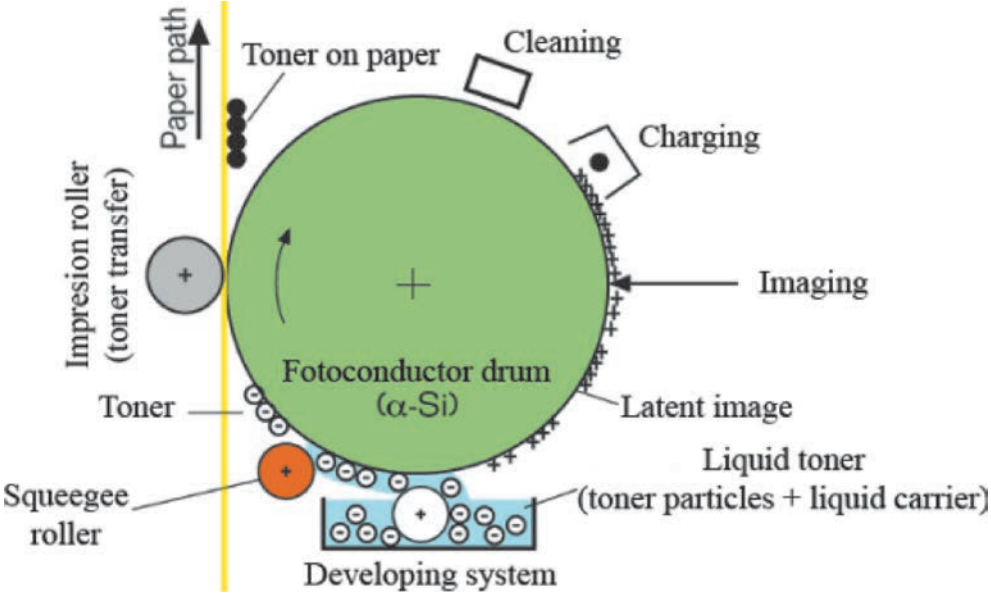


Fig. 1 – Indirect LEP technology (Majnarić, 2007)

The process of printing which uses the LEP technology has the following steps: charge the photoconductor drum, illumination the printing form, colouration the printing form with liquid toner, transfer the toner to the printing substrate, fix the toner on the printing substrate, and clean the printing form from residue ink.

This paper shows the results of the research conducted regarding the sample analysis related to the third, development stage in LEP technology, where pigment particles need be isolated from the liquid carrier by use of a squeegee roller. The squeegee roller reduces the total amount of liquid ink on photoconductor, thus creating a very thin layer on printing substrates. The photoconductor drum pressures the paper directly, thus obtaining high-quality print.

Phase of development is carried out in three separate parts: developing with the processing drum, developing with reverse rollers and fixing with squeegee roller.

Phase of developing is conducted in three separate parts: the developing with the processing drum, developing with reverse rollers and fixing with squeegee roller. All these aforementioned distinct parts of the developing phase are performed synchronized and they are the subject of research presented in this papar.

Machine calibration is followed by test printing:

1. Voltage of the processing drum: -200 V, -280V, -350V, -430V and -500V (series 1, samples: a_1, a_2, a_3, a_4 i a_5)

2. Voltage of the reverse rollers: 0V, -50V, -125V, -200V, and -250V (series 2, samples: b_1, b_2, b_3, b_4 i b_5)

3. Voltage of the squeegee roller: -1250 V, -1300V, -1350V. -1400V and -1450V (series 3, samples: c_1, c_2, c_3, c_4 i c_5)

During the experimental printing, only one parameter in the printing process was varied, while all other parameters remained constant as defined by the initial calibration of the printing machine.

The printing form contained different printing elements: a standard CMYK step wedge in the 10%-100% tone value range, a standard ISO illustration for visual control, textual positive and negative microelements, wedges for determining greyness and the standard wedge with 378 patches for the production of ICC profiles and a 3D gamut.

Liquid ElectroInk was used. It contains 5 % monomer pigment paste, 94 % highly volatile mineral oil and about 1 % agent for increasing electric conductivity (Landa

et al., 1988). The particles are 1-2 μm in size, dispersed in liquid medium, which results in a high-quality 2400 dpi print. The agents for increasing electric conductivity are highly polar molecules that receive the pigments to their positive side and turn toward less negative electrostatic field.

In LEP technology problem were with printing on paper because the liquid toner did not always adhere uniformly to it. It was used coated paper designed for printing on HP Indigo press that improve adhesion. This substrate has high brightness (ISO brightness = 93%).

For the recycling of the prints alkaline chemical deinking flotation was used. During the process the following chemicals were used: sodium hydroxide 1%, hydrogen peroxide 1%, sodium silicate 2%, DTPA 0.2% and a surface active substance 0.4 %. The consistency of the suspension in the disintegration phase is 10%, while in the flotation phase the consistency is 0.6%.

The handsheets were made from non-deinked and deinked pulp, with the usage of the Rapid-K□then sheet former, according to the standard ISO 5269 (ISO 5269-2, 2002). The following values were used for the measurement of the optical characteristics of laboratory handsheets from non-deinked and deinked pulp: the diffuse blue reflectance factor according to ISO 2470, effective residual ink concentration-ERIC according to TAPPI T 567 pm-97 (ISO 2470-1, 2009; TAPPI T 567 om-09, 2009).

A count of the residual dirt particles and area were assessed using Spec*Scan Apogee System image analysis software (TAPPI T 563-08/R, 2012). This system utilizes a scanner to digitalize an image. The threshold value (100), white level (75) and black level (65) were chosen after comparing the computer images to the handsheets.

2. Results and discussion

In order to determine the impact of the printing substrate and contribute to optimizing LEP technology print recycling, the research was conducted of the impact of voltage change in sample printing, throughout the development process (processing drum, reverse roller, squeegee roller), on the recycled fibre characteristics.

Figure 2 presents the total dirt count of handshets before and after the flotation of prints of each of the sample series for the lowest and highest negative voltage used in printing.

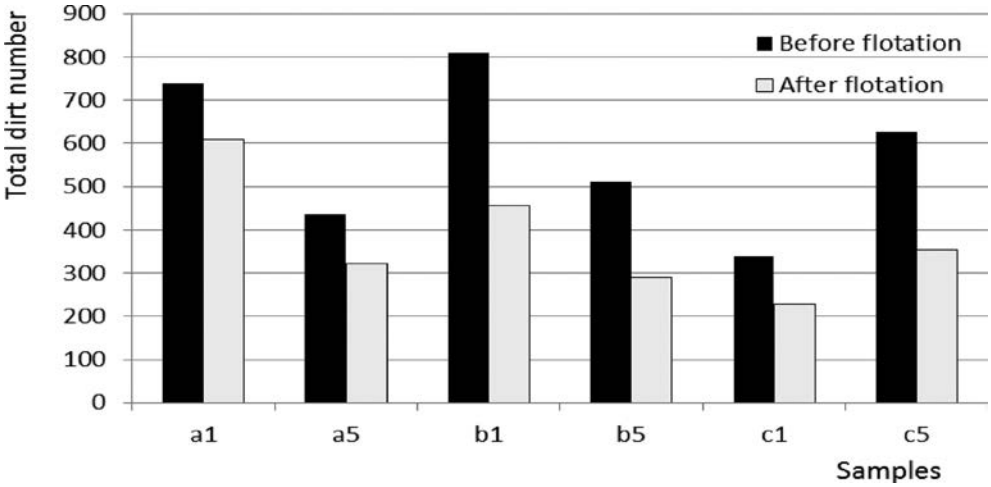


Fig. 2 – Total dirt count versus voltage of the processing drum, reverse roller and squeegee roller on handsheets from fibers after pulping and after flotation

Research results show that by increasing the negative voltage of the processing drum, and reverse roller the dirt count on handsheets resulting from the fibres after disintegration is reduced, while the results related to the squeegee roller are inversely proportional. The highest-level fragmentation is achieved with the lowest negative voltage prints of the reverse roller, followed by the lowest negative voltage prints of the processing drum (lower by 8.7 % compared to b_1) and the lowest negative voltage prints of the squeegee roller (lower by 58.3 % compared to b_1).

The lowest total dirt number is found on the handsheet made from recycled prints fibres with the lowest negative squeegee roller voltage (total number $_{b5}$= 227). The highest total dirt number is found on the sheets made from recycled fibres of the prints with the lowest negative voltage processing drum (total number $_{a1}$= 610). The lowest efficiency of removal the number of dirt by deinking flotation was achieved for the prints made at the lowest negative voltage processing drum (17.6 %), while the highest efficiency was achieved for the prints with the largest reverse roller voltage (46.2%).

Figure 3 shows the total dirt area of sheets obtained before and after the flotation of prints for the lowest and highest negative voltage used in printing.

The dirt takes the largest area on the handsheet made from fibres obtained by disintegrating prints made at the highest negative voltage of the reverse roller (b_5 = 119.875 mm^2). The difference versus other samples series being 75.8 % higher compared to prints b_1, 60.8 % higher compared to a_5 or 63.5 % higher compared to

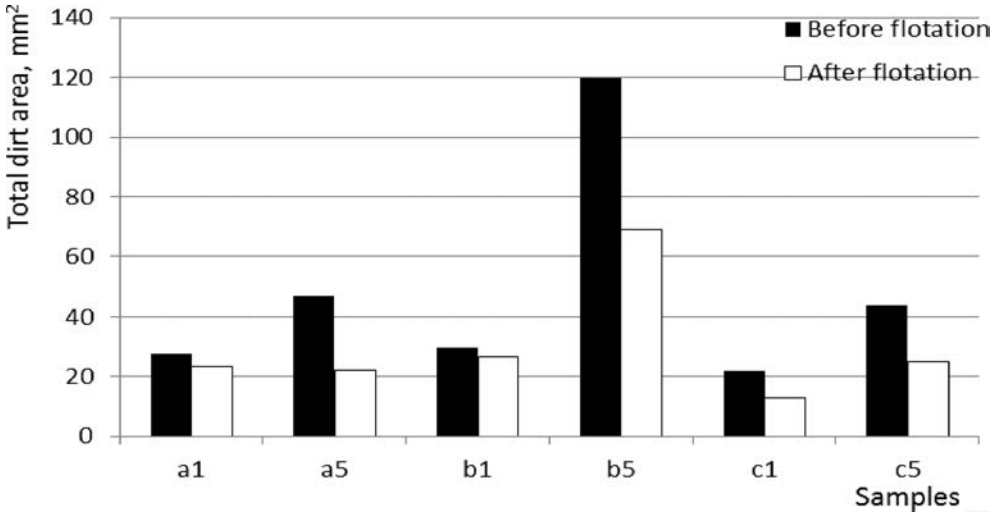


Fig. 3 – Total dirt area on handsheets from fibers after pulping versus voltage of the processing drum, reverse roller and squeegee roller

c_5. Furthermore, it should be noted that the area taken by dirt on the handsheet, as seen in Figure 3, increases with the increase in the negative voltage in the processing drum, reverse roller, and squeegee roller.

The lowest efficiency of removal the area taken by the dirt on the handsheet by the deinking flotation process was obtained for the prints made at the lowest negative voltage of the reverse roller (10.9 %), and the highest efficiency was achieved with the print at the lowest negative voltage of the squeegee roller (53.2%).

The flotation efficiency increases by prints with increasing negative voltages processing drum and a reverse roller (a_1 = 15.6%, a_5 = 52.8%; b_1 = 10.9%, b_5 = 42.3%), as opposed to the squeegee roller where the increasing negative voltage causes flotation efficiency deceases (c_1 = 41.3%, c_5 = 34.1%).

The results obtained indicate the need for a more detailed analysis of the dirt number and area, throughout the whole measuring range, as shown in Figure 4.

By increasing the negative voltage of the processing drum to the values of -200 V, -280 V, -350 V, -430 V, and -500 V, and of the reverse roller to the values of 0 V, -50 V, -125 V, -200 V, and -250 V, a decrease was observed in the dirt number on handsheets after pulping. However, the trend of decrease in the dirt count varies in monitored series. By increasing the negative voltage of the processing drum throughout the measured range, a decrease occurs in the dirt count, as follows: $\Delta\ a_2–a_1 = 125$, $\Delta\ a_3–a_1 = 213$, $\Delta\ a_4–a_1 = 315$ and $\Delta\ a_5–a_1 = 328$. By increasing the

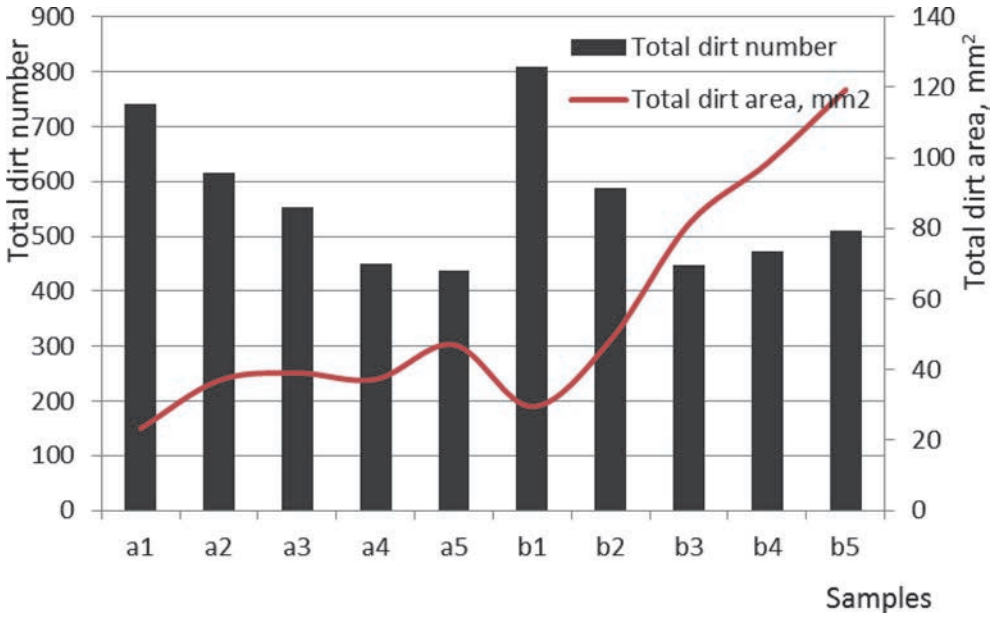


Fig. 4 – Dirt number and dirt area on handsheets from fibers after pulping

negative voltage of the reverse roller, a higher decrease in the dirt count occurs at lower negative values, compared to series 1 samples, while the values for samples b_4 and b_5 increased compared to b_3 ($\Delta\ b_2–b_1 = 221$, $\Delta\ b_3–b_1 = 362$, $\Delta\ b_4–b_1 = 337$, $\Delta\ b_5–b_1 = 358$).By increasing the negative voltage of the processing drum, in experimental conditions, the dirt area on handsheets after pulping is generally increased: $\Delta\ a_2–a_1 = 13.631\ mm^2$, $\Delta\ a_3–a_1 = 15{,}732\ mm^2$, $\Delta\ a_4–a_1 = 13{,}572\ mm^2$, $\Delta\ a_5–a_1 = 23.137\ mm^2$. A significantly higher increase of the dirt area is observed with handsheets made from fibres after pulping the prints obtained by an increase in negative voltage of the reverse roller: $\Delta\ b_2–b_1 = 18.311\ mm^2$, $\Delta\ b_3–b_1 = 51.712\ mm^2$, $\Delta\ b_4–b_1 = 68.911\ mm^2$ and $\Delta\ b_5–b_1 = 89{,}712\ mm^2$.

The results obtained may be explained by the principle of liquid toner development. The processing drum stands next to the photoconductor. Due to adversary rotation, the processing drum does not come into contact with the surface of the photoconductor. There is a very narrow gap between them into which the ElectroInk is injected. During printing, negative voltage is formed on the processing drum, creating electrostatic field with the photoconductor. In order for virtual printing elements to adhere the ink, they need to be more electropositive than the processing drum. With an increase in the negative voltage of the processing drum, the power of the electrostatic field increases, resulting in an increase in the application of ink on the printing element. Hence, the thickness of ink on the prints depends on the voltage of the processing drum.

The aim of the development process using the reverse roller is getting a clean photoconductor. Therefore, mineral oil is used on the surface of the photoconductor. An electrostatic field is formed in the contact zone between the photoconductor and the reverse roller, much weaker than the field formed between the photoconductor and the processing drum. Due to low voltage (0 V, -250 V), the reverse roller may easily become more electropositive than the printing elements, which results in additional detachment of ink from the photoconductor.A slight change in the voltage of the reverse roller affects the thickness of ink on the prints.

Varying the voltage of the squeegee roller may result in decreasing the thickness of ink on the photoconductor by over 50 %.

Since cases were observed in which the dirt number has little change, while at the same time the dirt area is significantly increased, as was the case with samples b_3, b_4 and b_5, the emergence of large, visible dirt can be expected in such cases, which can not be removed by flotation nor screening since they are too soft and pass through screens. This calls for a more detailed analysis of the dirt size, as shown in Figure 5.

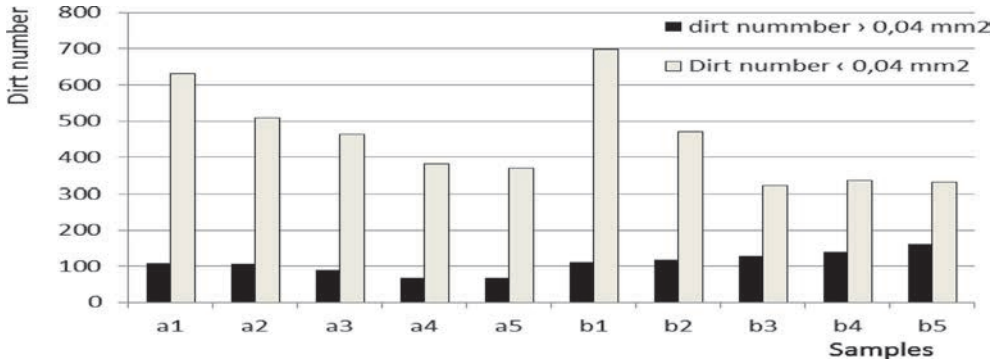


Fig. 5 – Dirt number > 0.04 mm^2 and < 0.04 mm^2 on handsheets from fibers after pulping

With the increase in negative voltage of the processing drum in experimental conditions, a decrease in the dirt number >0.04 mm^2 occurs on the handsheet made from fibres after pulping the prints, as follows: $\Delta\ a_1–a_2 = 3$, $\Delta\ a_1–a_3 = 19$, $\Delta\ a_1–a_4 = 40$ and $\Delta\ a_1–a_5 = 42$ and the dirt number < 0.04 mm^2 for $\Delta\ a_1–a_2 = 122$, $\Delta\ a_1–a_3 = 169$, $\Delta\ a_1–a_4 = 250$ and $\Delta\ a_1–a_5 = 261$.

The trends of sample series 2 differ from those of series 1. In this case the dirt number > 0.04 mm^2 is increased with an increase in the negative voltage of the reverse roller by the following amounts: $\Delta\ b_2–b_1 = 8$, $\Delta\ b_3–b_1 = 19$, $\Delta\ b_4–b_1 = 40$ and $\Delta\ b_5–b_1 = 42$. The dirt number < 0.04 mm^2 has a decreasing trend for b_1, b_2 and b_3,

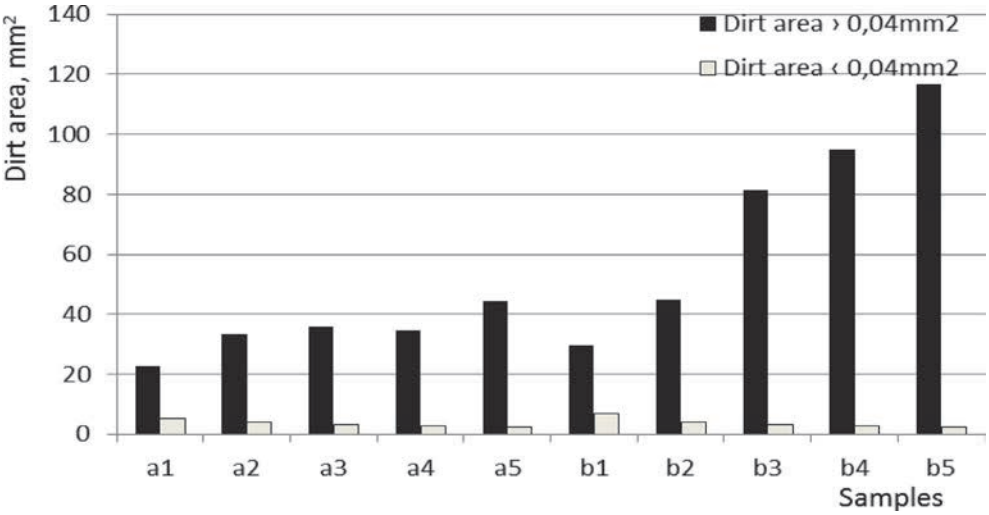


Fig. 6 – Area covered by dirt > 0.04 mm^2 and dirt < 0.04 mm^2 on handsheets from fibers after pulping

while b_4 and b_5 being slightly higher compared to b_3 ($\Delta b_1–b_2 = 229$, $\Delta b_1–b_3 = 378$, $\Delta b_1–b_4 = 364$ and $\Delta b_1–b_5 = 368$).

The area covered by dirt > 0.04 mm^2 on handsheets after pulping of series 1 with an increase of negative voltage of the processing drum shows a significantly lower increase ($\Delta a_2–a_1 = 10.512$ mm^2, $\Delta a_3–a_1 = 13.232$ mm^2, $\Delta a_4–a_1 = 12.032$ mm^2, $\Delta a_5–a_1 = 21.900$ mm^2) compared to series 2 ($\Delta b_2–b_1 = 20.511$ mm^2, $\Delta b_3–b_1 = 53.612$ mm^2, $\Delta b_4–b_1 = 72.111$ mm^2 and $\Delta b_5–b_1 = 93.941$ mm^2). The area taken by dirt < 0.04 mm^2 is considerably smaller compared to dirt > 0.04 mm^2, both for series 1 ($\Delta a_2–a_1 = 1.1712$ mm^2, $\Delta a_3–a_1 = 1.752$ mm^2, $\Delta a_4–a_1 = 2.350$ mm^2, $\Delta a_5–a_1 = 2.502$ mm^2) and for series 2 ($\Delta b_2–b_1 = 1.171$ mm^2, $\Delta b_3–b_1 = 1.901$ mm^2, $\Delta b_4–b_1 = 3.120$ mm^2 and $\Delta b_5–b_1 = 3.273$ mm^2).

In order to determine more precisely the number, size and area of dirt distribution was used dirt spot size from 0.001-0.006 mm^2 to >=5mm^2 . Figure 7 shows the number of dirt larger than >=5 mm^2 for all sample series.

The majority of large dirt >5 mm^2 forms with the increase in the negative voltage of the reverse roller, while the smallest amount forms with the increase of the negative voltage of the squeegee roller. One general feature is that at lower negative voltages all of the three print series observed no dirt size >= 5 mm^2 is formed. Large particles cause optical inhomogeneity of handsheets.

The criteria for cleanliness is a minimum dirt count. The effective residual ink concentration (ERIC) measurement means to relate ink content with brightness. An ink

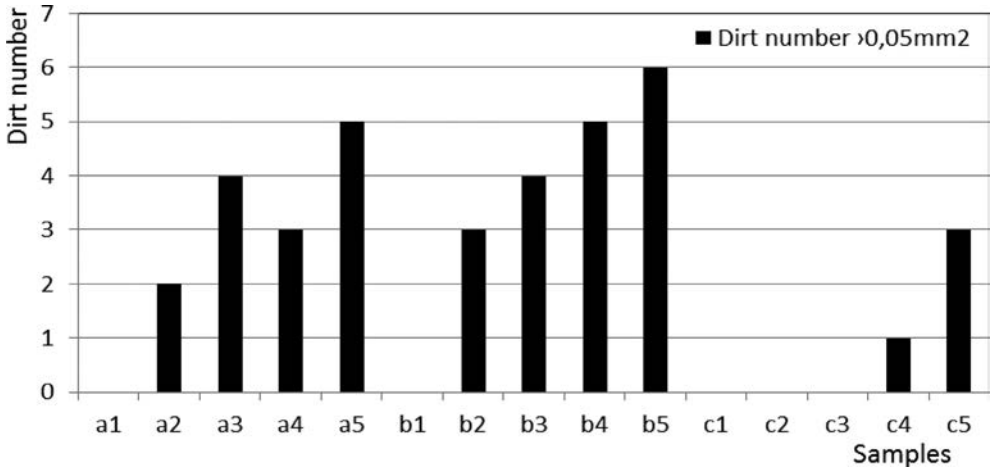


Fig. 7 – Dirt number >=5mm^2 on handsheets from fibers after pulping prints for series 1, 2 and 3

removal efficiency based on ERIC and dirt count across the different recycling operations tells what has happen to the ink, because uses refletance measurements in the infrared area of the spectrum where the absorption coefficient for the ink is several orders of magnitude greater than the absorption coefficient for the fiber, fillers and other components.

The results presented in Figure 8 show that the least number of remaining ink particles is found on handsheet made from deinked pulp of b_1 and b_5 prints. The largest difference in brightness was observed with the samples on handsheets before and after flotation (brightness gain b_1 = 2.0; brightness gain b_5 = 2.9).

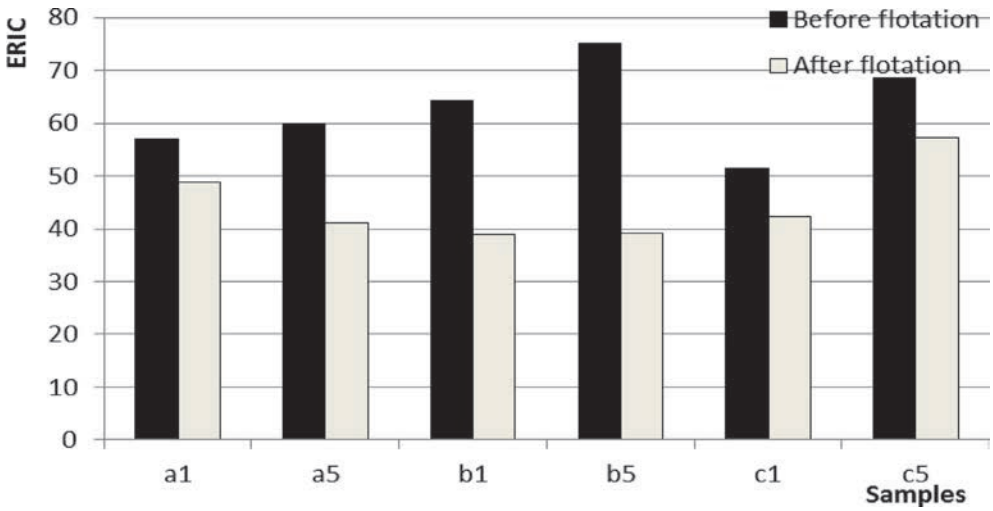


Fig. 8 – ERIC on handsheet from fibres before and after flotation

The differences in the values of effective residual in concentration on handsheets before and after the deinking flotation are good indicators of the efficiency of the recycling process and confirm the results obtained by other methods. The ERIC measurement is dependent on the distribution of ink particle sizes and is most effective for submicron particles.

3. Conclusion

The dependence of changes in the voltage of the processing drum, reverse roller and squeegee roller was determined in the LEP technology on the mechanism and efficiency of the deinking flotation process and characteristics of the recycled fibres.

During the flotation process, a slight formation of foam was observed, which is needed for successful dirt separation. This might be caused by the coating of the printing substrate, necessary to improve adhesion and increase the ink's adherence to the paper. In recycling process dispersants are surface active and together with alkali can lead to acceptable ink detachment from the coated paper. These species can hydrophilise ink containing agglomerates and hinder flotation efficiency as well as affecting foam stability.

Furthermore, the size of ElectroInk dirt also has an important role in the deinking of LEP prints – hence, by reducing dirt, deinking would be easier.

References

[1] Bhattacharyya, M, Manoj, K., Ng, H.,T., Mittelstadt, L., Laurie S., Aronhime M. (2011) Fatty acid based alkaline deinking of digital and non digital prints. Proceedings of 27th NIP International Conference on Digital Printing Technologies and Digital Fabrication, Minneapolis, 719-721

[2] Bhattacharyya, M, Ng, H.,T., Mittelstadt, L., S., Aronhime M. (2012) Effect of paper on LEP digital print deinking with alkaline and neutral chemistries. Proceedings of 28th NIP and Digital Fabrication, Quebec City, 526-529.

[3] Bolanaca Mirkovic, I., Bolanca, Z. (2005) Optical properties of deinked pulp. J. Imaging Sci. and Tech. 95(3)284-292

[4] Bolanaca Mirkovic, I., Bolanca, Z., Majnaric, I. (2014) Prints recycling in function of the press conditions and substrate characteristics. Procedia Engineering, 69(2014)150-157

[5] Bolanca Mirkovic, I., Majnaric, I., Bolanča, S., (2012) Recycling optimisation of the electrophotographic prints. Proceedings of the 8th International Conference of DAAAM, Otto, T., (Ed.), Baltic Industrial Engineering, Tallinn, 2012, 119-124.

[6] Bolanca Mirkovic, I., Majnaric, I., Bolanca, Z., Grgasović, A. (2009) Recycling of waste paper with different density of liquid toner. Proceedings of the 20th International DAAAM Symposium, Katalinić B. Ed. DAAAM International Vienna, 2009, 1411-1412.

[7] Bolanča, Z. (2000) Digital Printing and Environment. Annual of the Croatian Academy of Engineering, D. Aničić (Ed), Croatian Academy of Engineering, Zagreb, 71-74

[8] Bolanča, Z., Agić, D., Bauer, K. (2000) Recycling of the Digital Prints, Advances in Digital Printing. IARIGAI Conference, Montreal A. J. Bristow (Ed), PIRA International Ltd, Surrey, 81-84

[9] Carre, B., Mangin L. (2002) Digital printing – a threat to the deinking industry? PTS-CTP Symposium, Bordeaux, 7-2 to 7-15

[10] Carver, J., Guidry, N. (2011) Rethinking paper and ink: The sustainable publishing revolution. Ooligan Press, Oregon

[11] Fisher, A. (2008) Digital prints-not all of them are deinkable. INGEDE Seminar, Vienna

[12] Gmelin, H., Senring, S (2014) Determinations of a sustainable new product development. J. Clean. Prod. 154, 1-9.

[13] Heweltt-Packard Indigo, (2012) HP Indigo position steatement on deinking of HP Indigo ElectroInk prints. http://www. hp com. /go/indigo Accessed: 15. 05. 2015.

[14] INGEDE, (2008) Dry toners far better deinkable than liquid toner. Press release 3/2008. http://www.ingede.org/ingindxe/press/pr0803.html Accessed 15.05.2015

[15] INGEDE, Methode 11. (2001) Bewertung der Recyklierbarkeit von Druckerzeugnissen-Prüfung der Deinkbarkeit. INGEDE e. V. Bietigheim-Bissingen 1-7

[16] INGEDE, Methode 11. (2012) Bewertung der Recyklierbarkeit von Druckerzeugnissen-Prüfung der Deinkbarkeit. INGEDE e. V. Bietigheim-Bissingen 1-13

[17] ISO 2470-1 (2009) Paper, board and pulps – Measurement of diffuse blue reflectance factor. Indoor daylight conditions (ISO brightness)

[18] ISO 5269-2 (2002) Pulp-preparation of laboratory sheets for optical testing. Part 2. Rapid Köthern method

[19] Kadam, R. S., Evans, M. A., Rothenberg, S. (2009) A comparative study of the environmental aspects of lithographic and digital printing process. Rochester Institute of Technology, Rochester

[20] Landa B., Ben-Araham, P., Hall, J., Gibson, G. (1988), US Patent No.4794651

[21] Lasage P., Schoonenberg P. (2010) Comparing potential environmental impacts of printing on a Hewlett-Packard Indigo 7000 and a specific competitive sheet- fed offset press. Sylvatica, Maine.

[22] Macias, M., Lane , G., Miller, N., Belson, J., Ng, H. T. (2010) Deinking of HP digital commercial prints. Proceedings of PaperCon, Atlanta, http://www. hpl.hp.com Accessed: 15. 05. 2015.

[23] Majnarić I. (2007) Studij indirektne elektrofotografije. PhD Thesis, University of Zagreb Faculty of Graphic Arts, Zagreb

[24] Mittelstadt, L., Ng, H.,T., Bhattacharyya, M., Zhang, W., Hanson, E., (2010) High quality deinked pulps via alkaline-based HPMA deinking chemistry. Proceedings of 9th Research Forum on Recycling, TAPPI PEERS Conference, 2010. 92-100

[25] Ng, H .T., Bhattacharyya, M. K., Laurie S., Mittelstadt, L. (2010) Deinking and recycling HP digitah inks: From lab to pilot scale. Proceedings of 9th Research Forum on Recycling, TAPPI PEERS Conference, 2010. 101-110

[26] Ng, H .T., Bhattacharyya, M. K., Mittelstadt, L.S., Hanson, E .G., (2009) Deinking of HP digital commercial prints: Effect of chemicals and their loadings on deikability. Proceedings of NIP 25th International Conference on Digital Printing Technologies and Digital Fabrication, Louisville, Kentacky, 173-179

[27] Ng, H .T., Bhattacharyya, M.K, Mittelstadt, L., Hanson, E. (2010 a) Pilot scale recycling of HP indigo printed media and mixed office waste. Proceedings of 26th NIP International Conference on Digital Printing Technologies and Digital Fabrication, Austin, Texas, 88-91

[28] Putz, H. J. K. Blasins, E. Hanacker, (2008) Simulation of a two-loop deinking process in laboratory. Prog. Pap. Recycling, 18, 44-49.

[29] Schmitz, U., Fischer, A. (2004) Recycling killer in digital clothing: digital inks cause recycling problems. Druck Medien Magazine, Sept. 22-26

[30] TAPPI T 563-08/R (2012) Equivalent black area and count of visible dirt in pulp, paper and paperboard by image analysis.

[31] TAPPI T 567 om-09, (2009) Determination of effective residual ink concentration (ERIC) by infrared reflectance measurement.

[32] Wripp, M., (2010) Ingede declares Indigo print „unusable“. PrintWeek, 12. Nov. http://www.printweek.com/print_article/113042/ingede-declares-ind Accessed 12.05. 2015.

[33] Zhang, W., Bhattachatyya, M, Manoj, K., Mittelstadt, L., Ng, H.,T., Miller, T., Laurie, S., Aronhime, M. (2013) Laboratory scale two-loop deinking trials. Proceedings of 29th NIP International Conference on Digital Printing Technologies, Seatle, Washington, 2013, 422-424.

Professional paper

Implementation Details of National Identification and Authentication System

Martin Žagar*, Josip Knezović*

RIT Croatia, Faculty of Electronical Engineering and Computing

The e-government strategy of each country aims to deliver digital services to its citizens. In this paper Croatian National Identification and Authentication System (NIAS) as a base for usage of personal user box as a central point of receiving messages together with electronic identity infrastructure, as a key enablers in the development of trust into e-government in Croatia, will be desrcibed. Online authentication of electronic identity allows both the individual and the provider of government services (service agency) to have confidence in the identity of the other party in a transaction over the Internet (online transaction) and is essential infrastructural component for realizing overall objectives of e-government. There is a growing awareness that the interoperability of national public ICT infrastructures is a precondition for a more service-oriented and competitive public sector.

Key words:

National identification and authentication system, personal user box, electronic identity, digital business, e-government

1. Introduction

This paper decribes National Identification and Authentication System (NIAS) as a key enabling factor for development of user oriented public electronic services. Furthermore NIAS is responsible for authentication of entities which access common system and the exchange of identity information between entities that communicate with each other through a common system, or exchange documents and data

*martin.zagar@croatia.rit.edu, josip.knezovic@fer.hr

as well as verifying the authenticity of such identities. As a crucial benefit, NIAS provides a credible general framework of trust and identity management which greatly simplifies the necessary infrastructure, organization and services with significantly reduced costs in business.

In June 2010 Croatian Government started the procedure for defining the National Identification and Authentication System (NIAS) as a shared resource and building component of the national system to support interoperability among government entities involved in providing digital services. The action that followed was Croatian Government Decree on starting an e-Citizen project in April 2013. With this Decree basic public sector ICT infrastructure and framework for development of user oriented public services was set: Central government portal, National Identification and Authentication System and Personal User Box System. Similar concepts have already been realized in other countries worldwide as well. For example, Sweden has national intranet network for secure communication between government bodies and EU bodies as a part of e-Government service [1], Czech Republic offers citizens communication with the national authorities at one universal office, where you can receive or verify documents or acts from different institutions of public administration [2], Austrian Citizen Card can be used to sign documents electronically [3], Estonia first implemented X-Road infrastructure for cross-border services in domains not covered by existing EU and regional initiatives [4], Government Gateway in Great Britain enables people to communicate and make transactions with government from a single point of entry [5], VANguard is Australian government program that delivers cost-effective and reliable authentication services to secure business to government and government to government online transactions [6]. In Croatia, NIAS is designed on the principles of the EU project STORK (Secure Identity Across Borders Lined), respecting existing practices and accepted standards, so the electronic connectivity with EU member states can be established in the simplest and most effective possible way [7]. This paper describes the implementation details of central authentication and authorization system NIAS and use of personal user box for enabling digital resources of e-Government in the Republic of Croatia.

Figures 1 and 2 show increasing communication and interaction with public authorities in Croatia, although percentage is far below EU average. This shows raising confidence in e-government interactions in Croatia.

The rest of paper is organized as follows. In Chapter 2 initial registration to NIAS is described as well with simple user scenario, communication flow in providing of e-service is briefly described in Chapter 3, implementation results and discussion is provided in Chapter 4 with conclusion in Chapter 5.

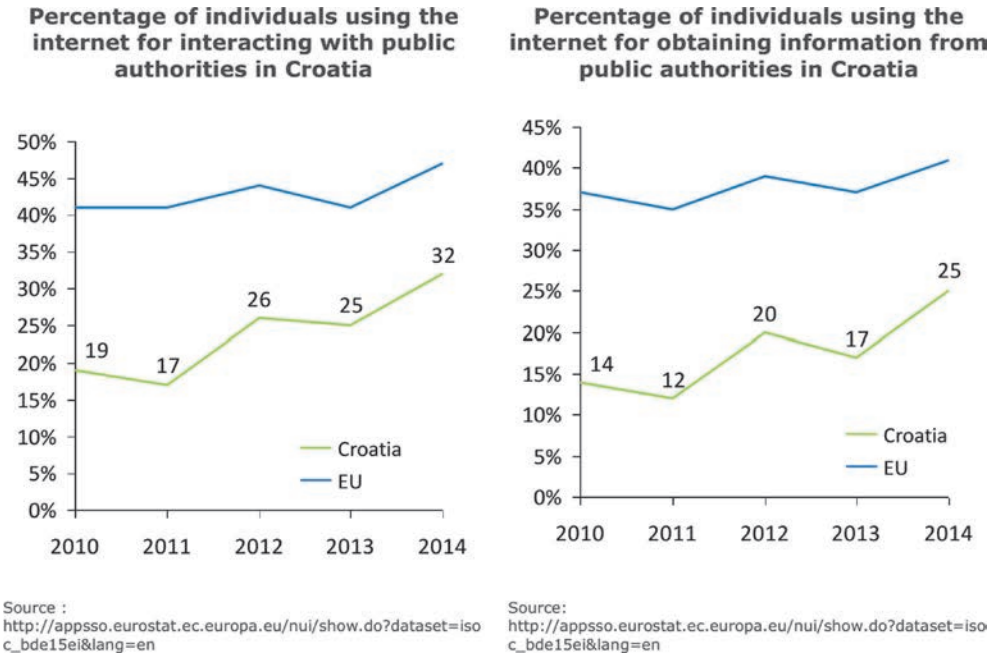


Fig. 1 – Percentage of individuals using the Internet for interacting with or obtaining information form public authorities in Croatia

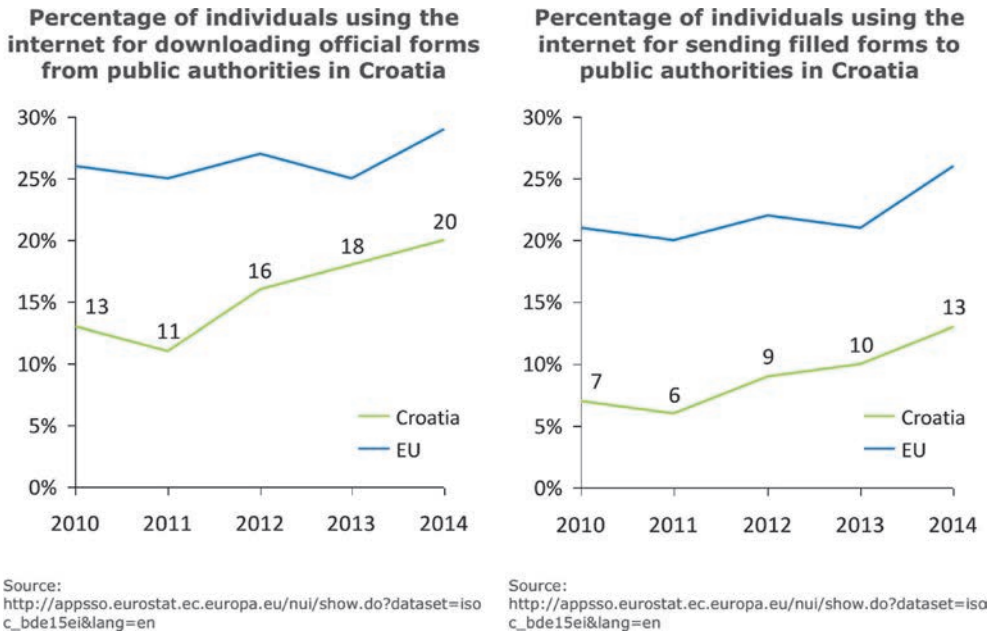


Fig. 2 Percentage of individuals using the Internet for downloading official forms or sending filled forms to public authorities in Croatia

2. Electronic Identity Registration and Operation

NIAS is designed with the goal to be flexible enough to allow interoperability and exchange of basic and specific attributes of the users of the system in order to provide electronic services (government bodies, local bodies, organizations) or define a relationship of trust in order to reduce unnecessary redundancy, i.e. increase the efficiency of the system. Figure 3 describes the scenario of the initial user registration in order to obtain electronic identity through NIAS.

As a part of Step 1, user applies to the registration procedure for electronic identity. The basic principle that should be guided by is availability of system for all users, and thus the wide availability of registration. If it is a registration for higher security levels, the process requires a higher degree of reliability and thus the physical presence and identifying of users and possibly signature. Registration sites are those of the relevant institutions (in particular post offices all over the country, Croatian Financial Agency desks, special registration places that are defined and verified within NIAS agencies).

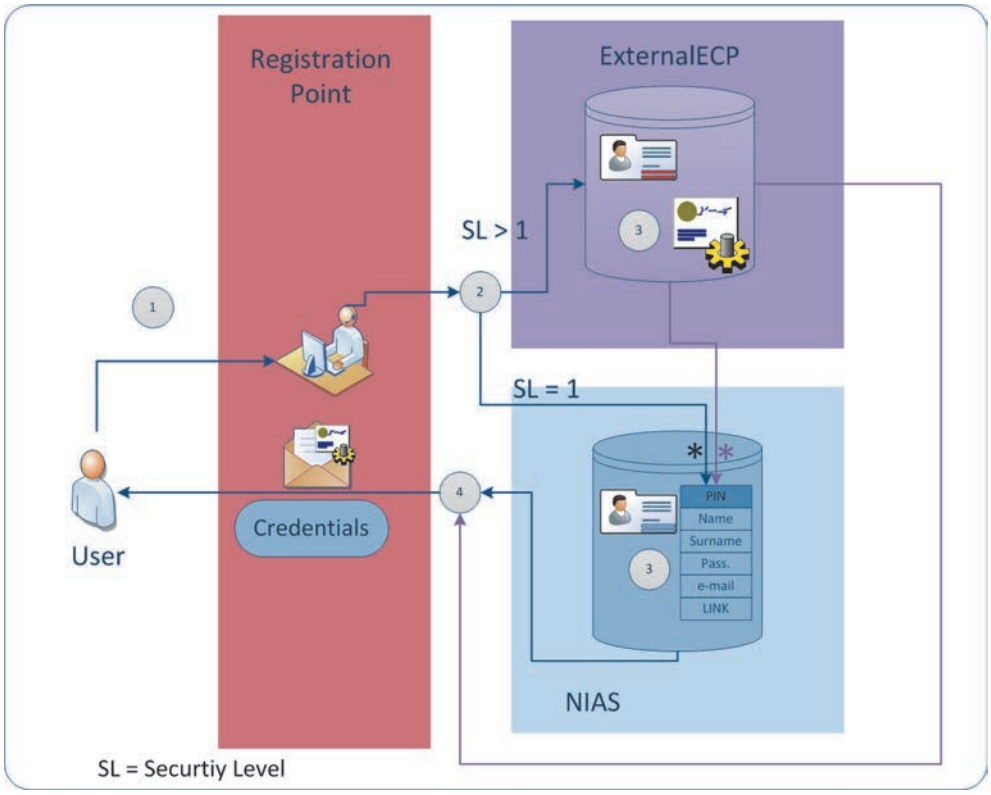


Fig. 3 – Initial registration

In Step 2, the application for registration is forwarded to external credential provider (ECP). For basic security level, the request is forwarded to the ECP component within NIAS that creates a basic set of data for electronic identity, while for higher level application is forwarded to the relevant external ECP (Step 3). In Step 4 ECP creates credentials and forwards them to the user (PIN/password, digital certificate, one-time-password (OTP), etc.). From this moment, the electronic identity of the user is active in the system. Information about the credentials or the electronic identity is recorded in the central NIAS directory. At this point, a basic role is assigned to the entity in the NIAS. Additional roles are subsequently assigned under local authorities if required, independently of NIAS.

Figure 4 depicts the user scenario of electronic identity in following Steps:

1. User (citizen, company representative, an official in the administrative body) accesses the unified entry point through the government portal in order to achieve e-services.

2. To prove his identity, request for verification of identity shall be forwarded to NIAS system.

3. If the required security level for the specified service level is equal to the security level 1 managed by NIAS, NIAS system will perform electronic verification of credentials. If the required security level is higher than 1, NIAS will request for verification of identity forward to outside ECP system who has a contract for such verification level.

4. The user proves his identity by forwarding his credentials to component for verification that is located within NIAS system in case of security level 1, or as part of the external ECP system for security level 2 or higher.

5. Verified identity is created as a result of Step 4.

6. Proof of verified identity is forwarded to authorization component of components service provider. From this point, the process of authentication is completed and begins an authorization procedure whose successful outcome will initiate the service.

7. Authorization component of service provider checks authority of electronic identity and obtained credentials. Assignment of rights is under jurisdiction of the body that provides a particular service.

8. User has consumed/not consumed electronic service.

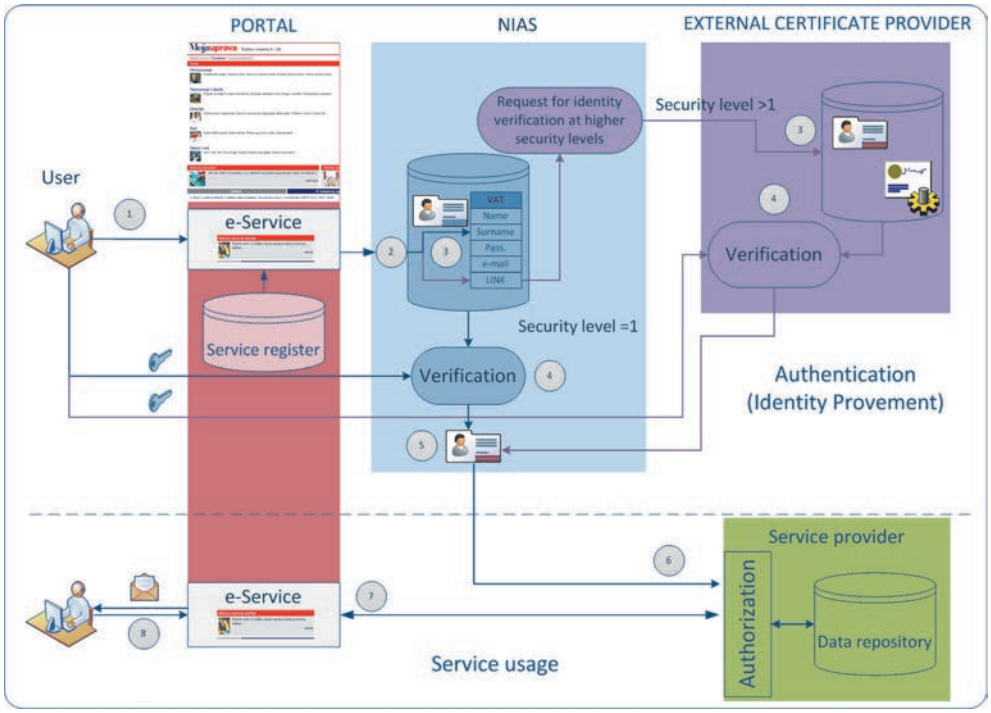


Fig. 4 – Use of electronic identity in the e-Government

3. Communication flow in providing of e-service

Figure 5 shows the flow of communication between user, service provider, NIAS and the authentication server. For the integration of electronic services in NIAS it is necessary that a service provider implements part of the SAML (Security Assertion Markup Language) protocol that sends a SAML request (SAMLRequest) for authentication (step 2 in the diagram) and receives SAML response (SAMLResponse) and processes it (step 9 in the diagram). For the integration of the credential issuer in NIAS it is necessary that the credential issuer on the authentication server implements part of the SAML protocol which enables receiving a SAML request (SAMLRequest) for authentication (step 5 in the diagram) and sending SAML response (SAMLResponse) to NIAS server (step 8 in the diagram).

In accordance with the Croatian Government Decree on starting an e-Citizen project, the electronic identity infrastructure and its carrier NIAS have been recognized as one of the key conditions necessary for the realization of a technologically advanced system of public administration and the inclusion of interoperable systems in the EU. Electronic identity as a primary objective of NIAS

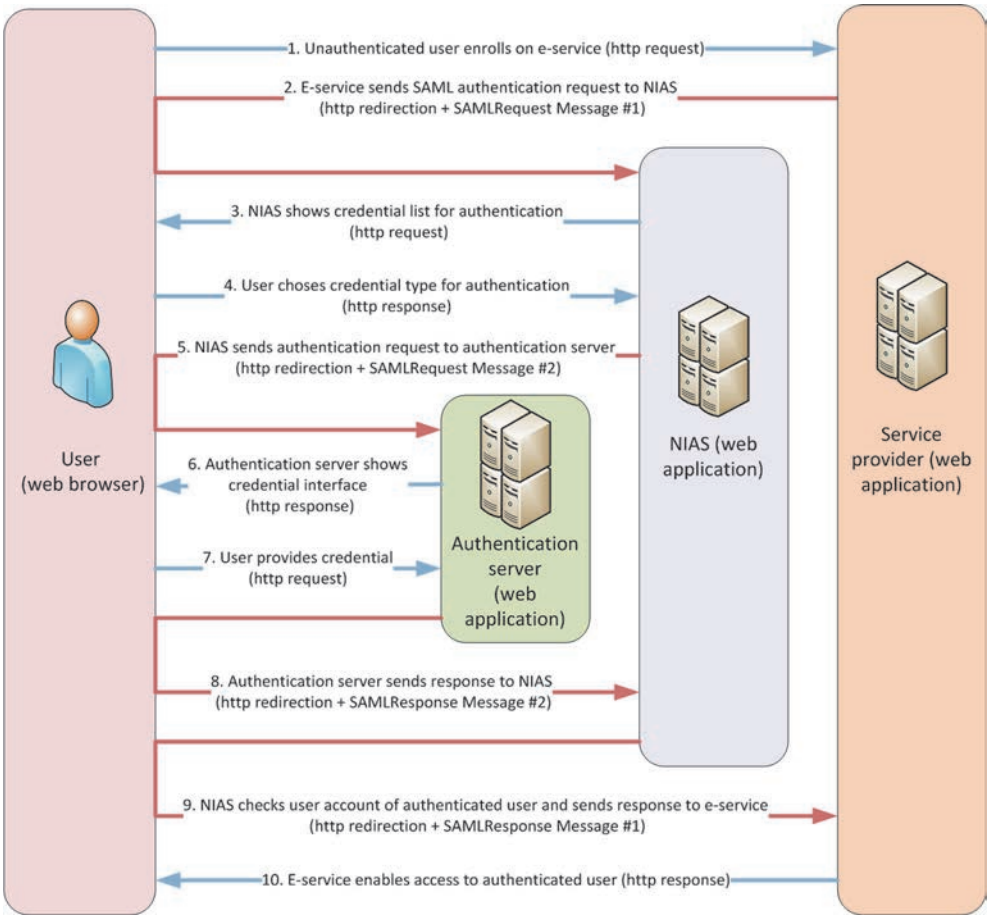


Fig. 5 – Communication flow in providing of e-service

was therefore recognized as a key strategic objective within the implementation strategies.

Additionally, since every public institution has been obliged to use central electronic identity system – NIAS, two more infrastructure components arose as a crucial to being implemented. Central government portal – a central web resource where all public administration information and other public content are integrated under the common visual and editorial standards. Second, as many electronic services is expected to be implemented in the following years and not every of them to be used from all users, a Personal User Box System has been created to enable personalization of the access to the electronic services.

4. Results and Discussion

Personal User Box, shown in Figure 6, has been recognized as a vital component of e–Citizen project.

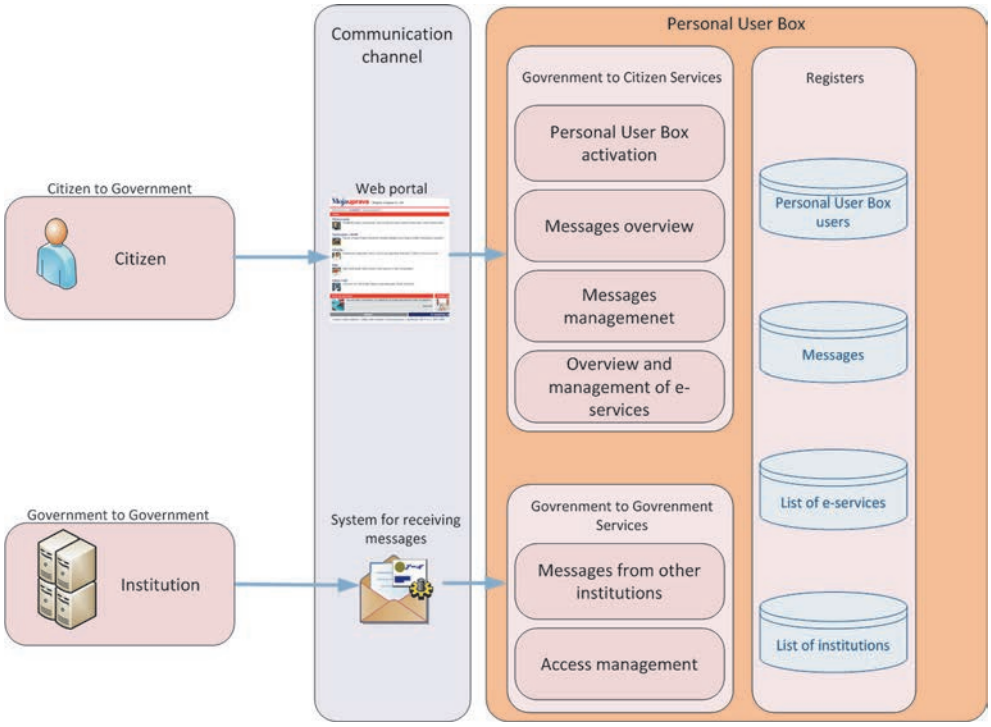


Fig. 6 – Personal User Box

The primary function is to receive messages that have been sent by the institutions/public authorities. The system for receiving messages has the function of formal and security control of messages received from institutions and forwarding messages in Personal User Box. Before start of sending the messages to Personal User Box, institution should request a registration in system for receiving messages, as well as the registration in Personal User Box system.

When institution creates message that should be sent to the Personal User Box it has to be created according to the agreed scheme defined by the Personal User Box system. The system is implemented as a web service (SOAP/HTTPS) and uses the Internet as a communication channel carried on HTTPS protocol for authentication of the communicating parties. The system supports the delivery of the message size up to 1MB.

5. Conclusions

Online government services enable citizens or representative of a business entity to access them at any time and from anywhere, regardless on working time or physical location of individual institutions. However, in such non secure electronic environments, it is necessary to provide a mechanism that will allow reliable identification of both parties in communication. This is accomplished in a way that each participant is given an appropriate electronic identity in the e-Government, assigned and guaranteed by the reliable component, NIAS, in which both parties (user and service provider) have full confidence.

NIAS system has the following advantages:

- Allows users to log into any online service using the same credentials, which may be a combination of a username/password, or a combination with a digital certificate, token or PIN
- Provides Single Sign On service throughout the system which allows users access to the resources and services at the present security level (within the Internet browser sessions), whereby the physical authentication and authorization is performed only once through a central system of authentication and authorization
- Allows users to use all services directly or delegate authorization to other government bodies with associated registers and systems applications
- Allows to connect user account with a wide range of different government identifications (for example, can be linked with health insurance number, personal identification number of academic citizens, and allows the use of these user attributes in the systems that require them
- Provides a safe environment and submission of electronic forms and transaction services to government bodies, using the same authentication credentials for signing submissions
- Provides a mechanism for secure communication between government bodies and users when exchanging of sensitive personal information
- Ensures interoperability in transactions between different services, regardless of specific applications or technologies used
- Defines the use of open standards in communication and technology neutral service
- Enables easier and faster way to connect to specific data registers on the Internet and routing of verified transactions between government bodies
- Reduces deployment costs, reduces risks with applying of well-known security protocols and reduces service delivery time
- The complexity of the business processes of the system is hidden to users, making it easier also for service providers and the delivery of services to customers.

References

[1] Swedish Government Secure Intranet, www.tutus.se/cases/sgsi.html (2015-06-29)

[2] Croatian Government Decree about NIAS, https://www.gov.hr/UserDocsImages/e-Gradjani_dok/2014-05-28-MUeH%20-%20Protokol%20rada%20NIAS-a%20-%20Ver.%201.3.pdf (2015-06-29)

[3] Mobile Phone Signature & Citizen Card, https://www.buergerkarte.at/en/ (2015-06-29)

[4] X-Road Europe, https://www.x-road.eu/about.html (2015-06-29)

[5] Government gateway, http://www.gateway.gov.uk/ (2015-06-29)

[6] United Nations E-Government Survey 2014, E-Government for the Future We Want, http://www.unpan.org/e-government (2015-06-29)

[7] STORK 2.0 project https://www.eid-stork.eu/ (2015-06-29)

[8] Adegboyega Ojo, Tomasz Janowski and Elsa Estevez, Determining Progress Towards e-Government – What are the Core Indicators?, UNU-IIST Report No. 360, United Nations Perceptions of transparency of government policymaking: A cross-national study, University International Institute for Software Technology, Macao, (2007)

[9] Saleh Alshomrani, A Comparative Study on United Nations E-Government Indicators between Saudi Arabia and USA, Journal of Emerging Trends in Computing and Information Sciences, Vol. 3, No. 3 (2012), pp. 411-420

[10] Jeannine E. Relly, and Meghna Sabharwal, Perceptions of transparency of government policymaking: A cross-national study, Government Information Quarterly, Vol. 26 (2009), pp. 148–157

[11] Rodrigo Sandoval-Almazan and J. Ramon Gil-Garcia, Towards an Evaluation Model for Open Government: A Preliminary Proposal, In proceedings: Electronic Government EGOV 2014(2014), pp. 47-59

[12] Enrico Francesconi, An Interoperability Approach for Enabling Access to e-Justice Systems across Europe, In proceedings: Electronic Government and the Information Systems Perspective EGOVIS 2014(2014), pp 26-41

[13] Shirish C. Srivastava, and Thompson S.H. Teo, The Relationship between E-Government and National Competitiveness: The Moderating Influence of Environmental Factors, Communications of the Association for Information Systems, Vol. 23 (2008), pp. 73-94

[14] José-Rodrigo Córdoba-Pachón, Patterns for E-Government Development, Systems Thinking and E-Participation: ICT in the Governance of Society, Kristin Klinger(Ed.), Information Science Reference, Hershey PA, (2010), pp. 33-55

[15] Donna Evans, and David C. Yen, E-Government: Evolving relationship of citizens and government, domestic, and international development, Government Information Quarterly Vol. 23 (2006), pp. 207–235

[16] Shirish C. Srivastava, and Thompson S.H. Teo, E-Government, E-Business, and National Economic Performance, Communications of the Association for Information Systems, Vol. 26 (2010), pp. 267-286

[17] Diego Navarra and Carmine Bianchi, Territorial Governance, E-Government and Sustainable Development Policy: A System Dynamics Approach, In proceedings: Electronic Government EGOV 2013(2013), pp 14-26

[18] Petter Gottschalk and Hans Solli-Sæther, Interoperability in E-Government: Stages of Growth, Integrating E-Business Models for Government Solutions: Citizen-Centric Service Oriented, Susheel Chhabra (Ed.), Information Science Reference, Hershey PA, (2009), pp. 50-67

[19] Member States' Competitiveness Report http://ec.europa.eu/growth/industry/competitiveness/reports/ms-competitiveness-report/index_en.htm (25-08-2014)

[20] EUROSTAT – Individuals using the internet for interacting with public authorities http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_bde15ei&lang=en (08-03-2016)

Scientific paper

Camouflage Military Uniform of the 21st Century

Đurđica Parac-Osterman[1,*], Martinia Ira Glogar[1,*], Ivana Žiljak Stanimirović[2,*]

[1]University of Zagreb Faculty of Textile – Technology, Croatia
[2]University of Zagreb Faculty of Graphic Arts, Croatia

In this paper, a part of the comprehensive research will be presented, which was performed as a part of a joint project of University of Zagreb faculty of Textile Technology and Croatian Ministry of Defence. One of the project aims wash defining the spectral characteristics of camouflage colours in VIS and NIR spectral area with retrospect on the role of design and new methodology based on macro/micro elements ratio in military camouflage pattern (versus currently adopted methodology of digital design). Also, the analyses of a spectral remission characteristics of chosen colours in compare to average shades of natural continental woodland environment have been performed over VIS and NIR spectrum in order to define them in a concept of camouflage effect achieving. In paper, also, a new technology INFRAREDESIGN® has been presented, which is based on resolving and controlling specific colorimetric characteristics and dyes properties in infrared spectrum. The technology processes hidden text, image or graphic, which is not visible by the eye in VIS spectrum but can be detected by using a new Z RGB camera only in NIR waveband. This technology is based on "Twin Dyes" preparation which have equal RGB values in VIS area but different Z values in NIR spectrum. The analyses of "Twin Dyes" performance in VIS and NIR spectrum has been performed.

Key words:

Camouflage, camouflage patterns, NIR/VIS, smart elements, INFRAREDESIGN® technology

[*]djparac@ttf.hr, martinia.glogar@ttf.hr, ivana.ziljak@grf.hr

1. Introduction

Systematic programs of contemporary military equipment development including highly specialized uniforms, started by the middle of the 1990's of a 20th century. High speed of technology and information sciences development enabled the development of highly sophisticated electronic and digital systems, which are implemented not only into a combat and technical equipment but, also, into a textile materials. Certain electronic elements, "smart" elements and elements of extended reality are implemented into materials aimed in highly functional military uniform production. In addition, with 21st century, there is a renewed interest in the design and evaluation of improved personal camouflage, implying satisfactory concealing performance of the military uniforms over the VIS and NIR spectrum. Successful blending with the natural environment (camouflaging) depends on a huge number of factors, but in general, it can be said that for an object to be successfully camouflaged it needs to closely match the environment's colours, shapes and textures. Personal camouflage (spotted uniforms), attempt to imitate the distribution of colours and edges in the visual environment providing a reflectance profile closely resembling its surroundings (Toet, A. et al. *Optical Engineering*, 2013., Rubeziene, V. et al. *Journal of Materials Science*, 2008, 2009).

Observation in the visual region (VIS) remains the primary means of military surveillance, however, modern battlefield surveillance devices may operate in one or more wavebands of electromagnetic spectrum. The emergence of SWIR (shortwave infrared) sensors and ongoing development of multi-spectral imagers that operate across visible, near infrared (NIR), SWIR and medium-wave infrared (MWIR), pose new challenges for current camouflage, concealment and deception concepts which are directed to mimic targets into operational background in all the relevant radiation spectral bands (Rubeziene, V et al. *Journal of Materials Science*, 2009., Baumbach, J., Fakin, D. et al. *The Proceedings of Colour- Effects and Affects*, 2008., Parac-Osterman, Đ. et al.*, Clothing and Design Conference*/Dubrovnik, 2008).

A major requirement of the systematic camouflage design is to obtain a colorimetric match of the concealed object to its anticipated surrounding, achieving the average reflectance of the material closely matched to that of the background across the required spectral regions. Specific interaction of chosen material characteristics, specified colour scheme and pattern as well as dyestuff properties, contribute to concealing properties of a specific product. From the aspect of the camouflage pattern, the key elements to provide satisfactory camouflage performance are shape, form, colour brightness and colour hue.

Traditionally camouflage design was largely based on intuition and aesthetics, and was performed by industrial designers and, especially in the beginning, by artists. The design principles were often inspired by nature and based on biological principles such as blending and disruption. By the end of the 20th century, scientific studies based on statistics of natural images, texture, visual perception and psychology have entered the design process. As a result, it is now commonly acknowledged that effective camouflage patterns should consist of both micro and macro-patterns and should be similar in composition to natural images. Current camouflage design follows two different trends: one trend is towards "universal" designs which perform well in a wide range of environments and another is towards more specialized designs and patterns with definite requirements for colour characteristics (Rubeziene, V. et al. *Journal of Materials Science*, 2008., Glogar, M. I. et al. *Proceedings of XIIIth International Izmir Textile and Apparel Symposium*, Izmir 2014., Glogar, M. I. et al.; *Book of Proceedings of the 8th Central European Conference on Fiber – Grade Polymers, Chemical Fibers* 2015, Zagreb, Marr, D. et al. *Technology Review*, 1978.).

The implementation of, so called, "smart elements" and "elements of extended reality" is also characteristic of the military clothing of 21st century. One of the recent technologies that uses the elements of extended reality and which become of a great interests for Croatian Ministry of defence and Croatian army, is INFRAREDESIGN® technology patented by group of scientists led by Vilko Žiljak, in 2008, (Ziljak, I. et al., Tekstil, 2009). INFRAREDESIGN® theory has developed a method of implementing graphics that carry two pieces of information, where the first one is seen in the visual spectrum (VS), and the other in the near infrared (NIR) spectrum. This technology of implementing hidden text, image or graphic which is not available to the human naked eye but visible only in near IR spectrum, into a camouflage pattern, allows marking and tagging individuals so as to recognize one's own forces in battle activities, in unfavourable meteorological conditions or during rapid movements of formations in complex situations. Hidden information is detected by new Z RGB camera (RGB which registers the RGB /red, green, blue/ in the visible light from 400 to 700 nm, and Z camera, which registers the NIR and measures the absorption of light in the Z size of 1000 nm). For Infraredesign theory (Ziljak, V. et al., The Imaging Science Journal, 2010.) special Twins dyes are projected and they have good absorption properties of a wider spectral area: VIS and NIR spectrum. Since these dyes include also Z size (Ziljak, V. et al., Technical Gazette, 2011), such dyes are called Z-dyes, and allow reading of hidden messages (Pap, K. et al., Journal of Imaging Science and Technology, 2010., Ziljak, V. et al., Infrared Physics &Technology, 2012).

In the context of current trends in military camouflage uniform development and production, as a part of cooperative project of University of Zagreb Faculty of Textile Technology with Croatian Army and Croatian Ministry of Defence, comprehen-

sive research has been performed, aiming in defining a proper colour scheme with definite spectral characteristics and remission properties, suitable for woodland terrain. For that purpose two original designs were created using new methodology of specific micro and macro-patterns partitions, obtaining composition in accordance to natural images. It is important to mention that since the 1970s, the digital camouflage patterns (camouflage pattern with pixelated look) have become widespread, but in time, certain disadvantages starts to appear. So parallel to overall usage of digital pattern, certain new methodologies for designing efficient camouflage patterns have been developed, such as methodology of multiscale patterns with specific partition of micro and macro elements without pixelated structure, or fractal based patterns which will provide satisfactory camouflage over a range of distances, meaning over a range of scales.

Based on such consideration, as already being mentioned, two original designs for the project of Croatian Ministry of Defence "New Croatian Camouflage Pattern", were created. The specific design used stylized shapes of animals (marten and griffon vulture) that are symbols of natural heritage of Republic Croatia and which are composed with other macro and micro elements of design into a pattern with concealing ability. Those specific patterns will be presented and analysed from the aspect of the form, shape and colour properties in the following chapters.

Also, a new INFRAREDESIGN® technology will be presented. The process of implementation of specifically engineered infrared message which is seen isolated at 1000 nm message into a camouflage pattern, and a process of creating specialized "Twin" dyes will be explained.

2. Methods

In the first part of the experimental work, the process of creating a two original patterns with specific composition of macro and micro elements will be presented. The patterns were created using stylized shapes of "marten" and "griffon vulture", and in a first step of the work those were created using a traditional methodology of a single baseline patterns. The patterns are shown on Figure 1.

The patterns were developed as an individual design consisting of larger number of macro elements and 4 standard camouflage colours, with colour characteristics that correspond to Croatian army technical requirements for woodland outdoor uniforms: CIE (D65/10) colour values; L*, C*, h, a* and b* (multi-coloured shade), Table 1. The instrumental colour measurement over VIS waveband range, with aim of CIE (D65/10) colour values obtaining, has been performed by means of remis-

Fig. 1 – Sample 1a – baseline "marten" pattern; sample 2a – baseline "griffon vulture" pattern

sion spectrophotometer DataColor® type SF600+CT, with constant instrument aperture (apertures "L"=2.6), D65, using d/8° geometry. Colour measurement in NIR waveband range has been performed by means of Specord 50 Plus BU spectrophotometer.

Table 1 – Croatian army colour characteristic requirements for woodland outdoor uniforms

Colour	*L**	*a**	*b**	*C**	*h*
Beige	44,05	1,96	16,45	16,57	83,2
Black	18,48	0,11	-1,17	1,17	75,62
Brown	28,75	7,15	12,24	14,17	59,72
Green	36,47	-8,21	12,75	14,61	119,22

The analyses of remission characteristics of chosen colours, in both VIS and NIR waveband have been performed in compare to the reflectance spectrum of the natural background, and are shown on Figure 2.

Reflectance spectrums of chosen colours are also in accordance to NATO standards. Figure 3 shows where NATO expects for camouflage colours reflectance to fall within the infrared camouflage, and it is suitable for certain range of woodland topography.

Based on analyses of natural background of Croatian continental woodland terrain (grass, leafs, bark and soil), following partitions of each colour used in camouflage patterns, Figure 1., have been determined: green 30%; brown 35%; beige 20% and black 15% (used for achieving imitation of shadows in nature).

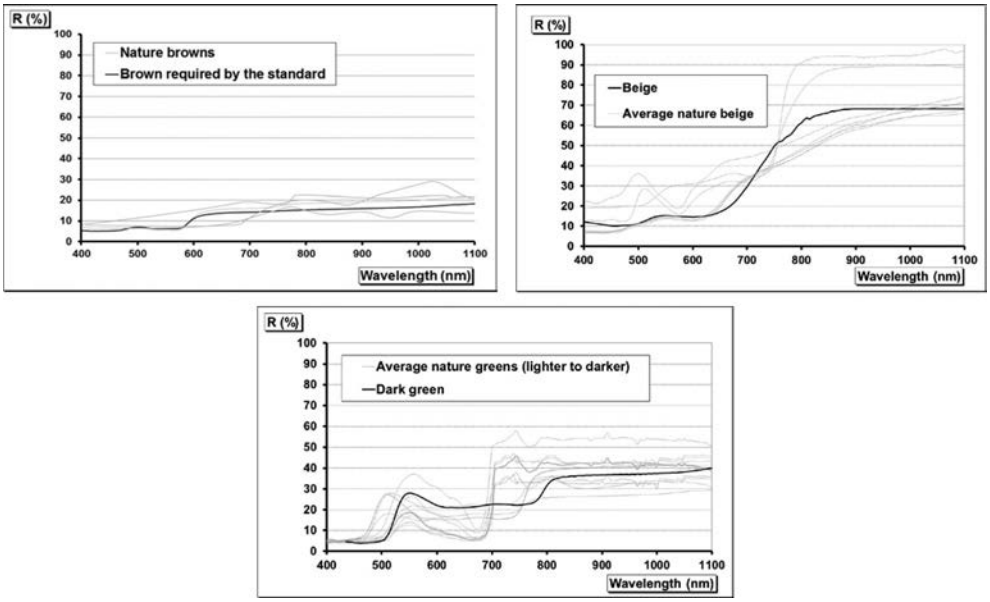


Fig. 2 – Reflectance spectrums of chosen brown, dark green and beige in compare to average spectrums of colours characteristic for Croatian continental woodland surrounding

In the further work two experimental variations of the original design were created and are shown on Figures 4 (Samples 1b and 1c) and 5 (Samples 2b and 2c). Figure 4 (Samples 1b and 1c) shows the variation of baseline pattern with pixelated structure (digital pattern).

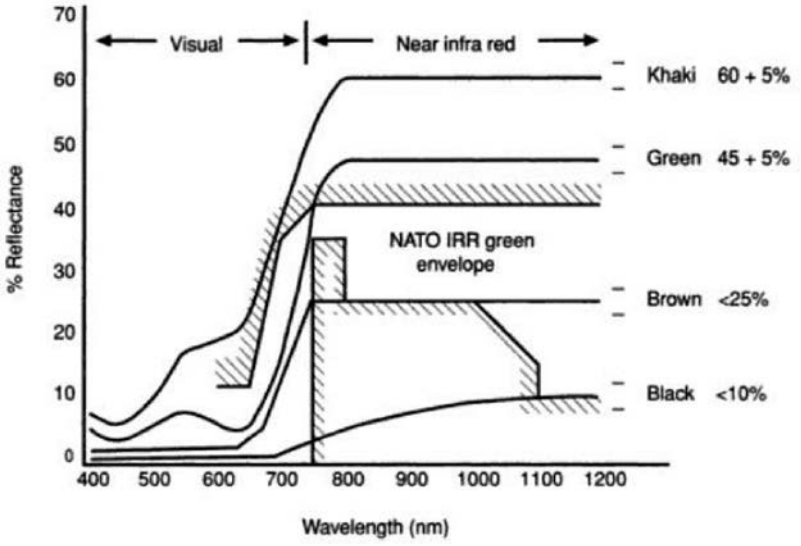


Fig. 3 – NATO requirements for camouflage colours spectral characteristics (Boyd, R. J., 2014.)

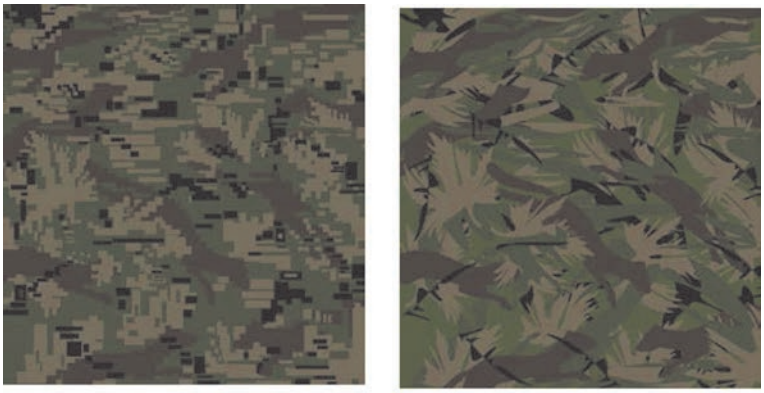


Fig. 4 – Pattern "marten": 1b – digital pattern; 1c – variation with larger number of micro elements and additional colour

Figure 5 (Samples 2b and 2c) shows the experimental pattern created with a certain partition of macro and micro elements with additional colour – light green and additional design element – stylized shape of an oak tree leaf, also characteristic for the Croatian natural heritage.

In the cooperation with "Čateks d.o.o.", Croatian textile company specialized in production of military programme, digitalized pattern 2b has been realised by the technique of textile rotary screen printing and were used for experimental military uniform production. The uniform was tested during one of the field examination performed by the Croatian army and the photo of uniform performance in VIS spectrum during daytime is presented on Figure 7, and explained in the section 3. Results and discussion.

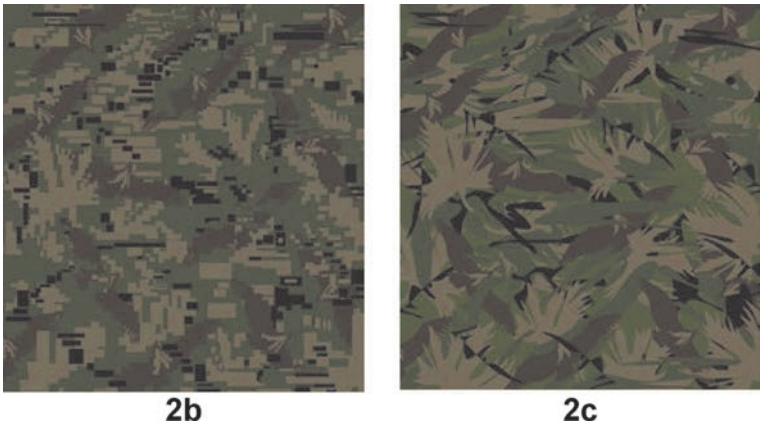


Fig. 5 – Pattern "griffon vulture": 2b – digital pattern; 2c – variation with larger number of micro elements and additional colour

Colour characteristic of new additional, light green colour used in patterns 1c and 2c, are shown in Table 2 in terms of CIE (D65/10) colour parameter values. Also, the remission spectra of chosen light green colour is given on Figure 6, in compare to average green spectrum characteristic for natural woodland surrounding.

Table 2 – Colour characteristics of additional, light green colour, according to CIE (D65/10) system

Colour	*L**	*a**	*b**	*C**	*h*
Light Green	40,67	-10,57	22,13	24,53	115,52

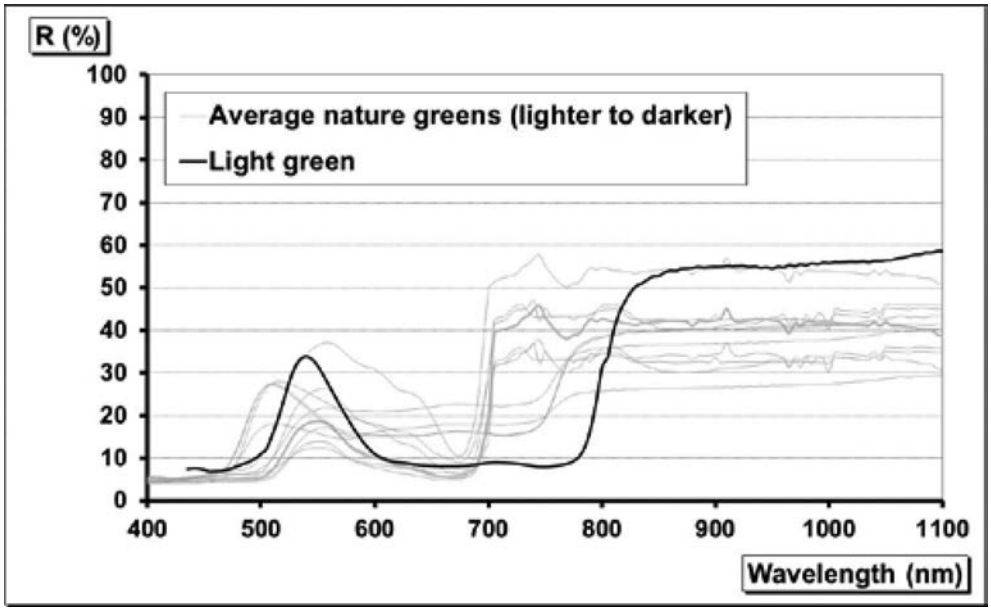


Fig. 6 – Reflectance spectrums of light green in compare to average spectrums of greens characteristic for Croatian continental woodland surrounding

By the method of photo modelling, the patterns *1a,b,c* and *2a,b,c* were interpolated into a natural scene of Croatian original woodland environment, in order to be tested and compared (Figure 8a and 8b).

In the following part of the research, some aspects of the INFRAREDESIGN® technology has been applied. Design 2c which was confirmed with more satisfactory concealing properties was implemented into a military pattern with hidden message (the Republic of Croatia coat of arms), which was printed on textile fabric using inkjet pigment printing technology (Hewlett packard designjet HP 5000) (Figure 10a and 10b).

For that purpose special "Twin dyes" has been created and analysed, which are the basis of camouflage painting which includes both hiding and showing of information. The spectral characteristics of selected "Twin dyes" are shown on Figures 9a – e.

Results and Discussion

Baseline patterns 1a and 2a (Figure 1), has been created in tradition military stile, with major partition of macro elements and in conventional 4 colour scheme for continental woodland terrain (green, beige, brown and black). Regarding the recommended NIR reflectance values for camouflage shades against IR radiation (700 to 1200 nm), which are, depending on the shade, as follows: for green 40-60%, for beige 55 – 65%, for brown and olive 20 – 40% and for black 10 – 20%, it can be seen on Figure 2, that remission spectrums of chosen colour scheme is in the recommended range for VIS and NIR waveband range, as well as in accordance to average spectrums of natural woodland environment. Characteristic design elements used for patterns 1a and 2a (Figure 1) were stylized forms of marten and griffon vulture, the characteristic animals which are parts of Croatian natural heritage and which represent one specific distinguishing element in pattern composition.

Variations 1b (Figure 4) and 2b (Figure 5) were created as digital versions of baseline patterns 1a and 2a (Figure 1), with pixelated contour. As it was explained in experimental part, although the digitalized pattern was created simply as version of baseline sample with pixelated contour as the only distinguishing element, during the joint project of University of Zagreb Faculty of Textile Technology and Croatian Ministry of Defence with the aim of designing a unique national military uniform, the sample 1b (sample with stylized shape of marten with pixelated form) was chosen for field testing and was realised in experimental military uniform production. The uniform is shown on Figure 7, blended in natural environment of Croatian continental woodland surrounding.

Fig. 7 – Military uniform produced using pattern 1b

During the subjective evaluation performed by the military trained observers it was determined that the overall concealing properties of the pattern regarding the shape, colour scale and form were satisfactory. But the fact that pixelated structure of the pattern is nowadays slowly replaced by the new methodologies of pattern composition based on certain proportion of micro and macro elements which ensures contrasts closer to nature, inspired further experimentation and as the results the patterns 1c (Figure 4) and 2c (Figure 5) were created.

Patterns 1 c (Figure 4) and 2c (Figure 5) were designed in a same form as 1a and 2a patterns (Figure 1), but with one specific design element added (oak tree leaf also as a part of Croatian natural heritage) and one additional colour (light green) included. This resulted in static pattern structure defined in original 1a and 2a to be broken into a smaller design elements which provided optimal partition of macro and micro elements. Dynamic form has been obtained which imitate vibrations, movement and disperse appearance closer to natural images. The most important characteristic of the natural woodland terrain is a presence of a multiple different scales of greens, disperse appearance and constant movement of different coloured elements, which can be more satisfactory achieved by the methodology of micro and macro elements proportion rather than by pixelated form which creates unnatural contrasts.

Also, another new element used in creating patterns 1c and 2c was a partition of each colour. In patterns 1c and 2c the partitions of beige and green are enhanced in compare to recommended, for woodland patterns. The values of partitions used are 23 % of beige and 41 % of green (dark and light green together). Also, the partitions of browns and black are decreased in compare to recommended, so the partition of brown is 30 % and black is 8 %. Those deviations are made in order to achieve dynamic composition of elements and their contrast relations adjusted to specific spectral characteristic for woodland environment. For black it is important to be in decreasing representation due to a fact that it is highly visible both to the naked eye and to modern optics. When viewed through night vision goggles, can appear excessively dark and create an undesirable high-contrast image.

By the method of photo modelling, the patterns created were interpolated into a natural scene of Croatian continental woodland terrain, in order to compare their subjective concealing properties (Figure 8a and 8b).

Patterns were tested for the ground terrain environment as well as for the tree tops and leaves environment. It can be seen that in general, patterns 1c and 2c, which were created on principles of new methodology of macro and micro elements partitions with one design element and extra colour added, exhibit results of more satisfactory incorporation into a natural surroundings. Patterns 1b and 2b also ob-

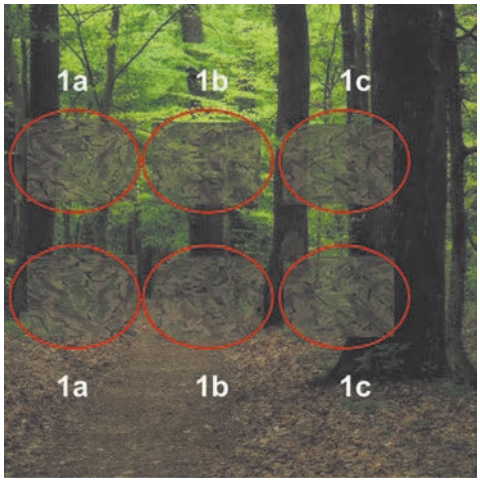


Fig. 8a – Interpolation of patterns 1a,b,c in natural scene

Fig. 8b – Interpolation of patterns 2a,b,c in natural scene

tained more satisfactory concealing appearance in compare to traditional baseline patterns, 1a and 2a.

In following work, the concept of INFRAREDESIGN® technology is presented (Marr, D. et al., Ziljak. V. et al., Pap K. et al.). The technology allows not only good camouflage of military uniforms, facilities, equipment, etc., but it processes also hidden information which are not visible in VIS spectrum. This technology of implementing hidden messages into a camouflage pattern, which is visible only in near IR spectrum, allows marking and tagging individuals so as to recognize one's own forces in battle activities, in unfavourable meteorological conditions or during rapid movements of formations in complex situations. INFRAREDESIGN® technology of production a camouflage patterns with hidden messages implies creating dyes with camouflage properties with same reflectance characteristics for observer's eye (VIS spectrum) but with different response in the near-infrared (NIR) waveband. Such dyes that have the same RGB values for visual (V) spectrum and different Z values for near infrared (NIR) spectrum are called "Twins". The value *Z* is the numerical value of absorption in the NIR-Z spectrum and it is introduced into standard colorimetric defined by three parameters (CIELab, RGB, HSB) as the fourth value which describes material light absorption in the sub-area of the near infrared (NIR) spectrum at 1000 nm.

To realize the idea of creating a camouflage pattern with hidden message, the design 2c, presented in the previous chapters of this paper, has been chosen. Besides demonstrating the colour properties in order to achieve compatibility with spectral characteristics of nature, in visible and NIR spectrum, a specific hidden message

(official Croatian Coat of Arms, Figure 10b), has been implemented into existing pattern and was printed on textile fabric using inkjet printing technology (Hewlett Packard Designjet HP 5000).

The first step in a process of creating a pattern with hidden message was "Twin" colour processing. The Twin dyes should have a match of a remission curves thorough visible part of the spectrum, while in NIR spectrum one of them should have no response at 1000 nm.

It is important to mention that the "Twin" dyes processing is a long, complex iterative process because each material with a specified dyes and printing techniques gives its own results. With the development of INFRAREDESIGN® (IRD®) methods and application to different materials, a need of extensive experimental work has been performed, in which for an each given material and dye an unique IRD® solution is achieved; in which the twins dyes in the visual part of the spectrum V (400-700 nm) match, but they respond differently to IR in the value of Z (1000 nm). The same group of dyes can vary drastically on various materials. For an each combination of printing, dyes and printing material iterations must be made with an aim of achieving the equal value of twins.

On Figures 9a to 9e the spectral characteristics that describing the properties of "Twin" dyes, are shown through remission plots of colours used in pattern with hidden message. Based on similarity of remission curves of "Twin" dyes, it can be expected that the total colour difference (dE* (D65/10)) will be in ranges of tolerances, confirming the properties of equality of "Twin" dyes in visible spectrum. Remission plots (Figures 9a – e), are showing two remission curves in wavelength range from 400 to 1000 nm (VIS and NIR spectrum), Twin 1 assigned as Z_0 and Twin 2 assigned as Z_{40}. Z_0 ("Twin 1") dye have no response at wavelength of 1000 nm (NIR spectrum), while Z_{40} ("Twin 2") dye is hidden Twin with equal properties in visible spectrum but with response at wavelength of 1000 nm (NIR spectrum).

"Twin 1" dye is produced from the process colours CMY, while "Twin 2" dye is produced by adding also K black colour (K). In all "Twin" dyes the partition of K colour is 40% assuring the equal value of coverage. Z_{40} value determines the equality of hidden (Z) graphic appearance for any hue applied in camouflage pattern with hidden message production. Those measurements are given for the real print outs on textile fabric. For each printing technique, material and dye, it is necessary to calculate new colour settings.

Based on similarity of remission curves of "Twin" dyes (Figure 9a – e), it can be expected that the total colour difference (dE* (D65/10)) will be in ranges of tolerances (DE<2,1), confirming the properties of equality of "Twin" dyes in visible spectrum.

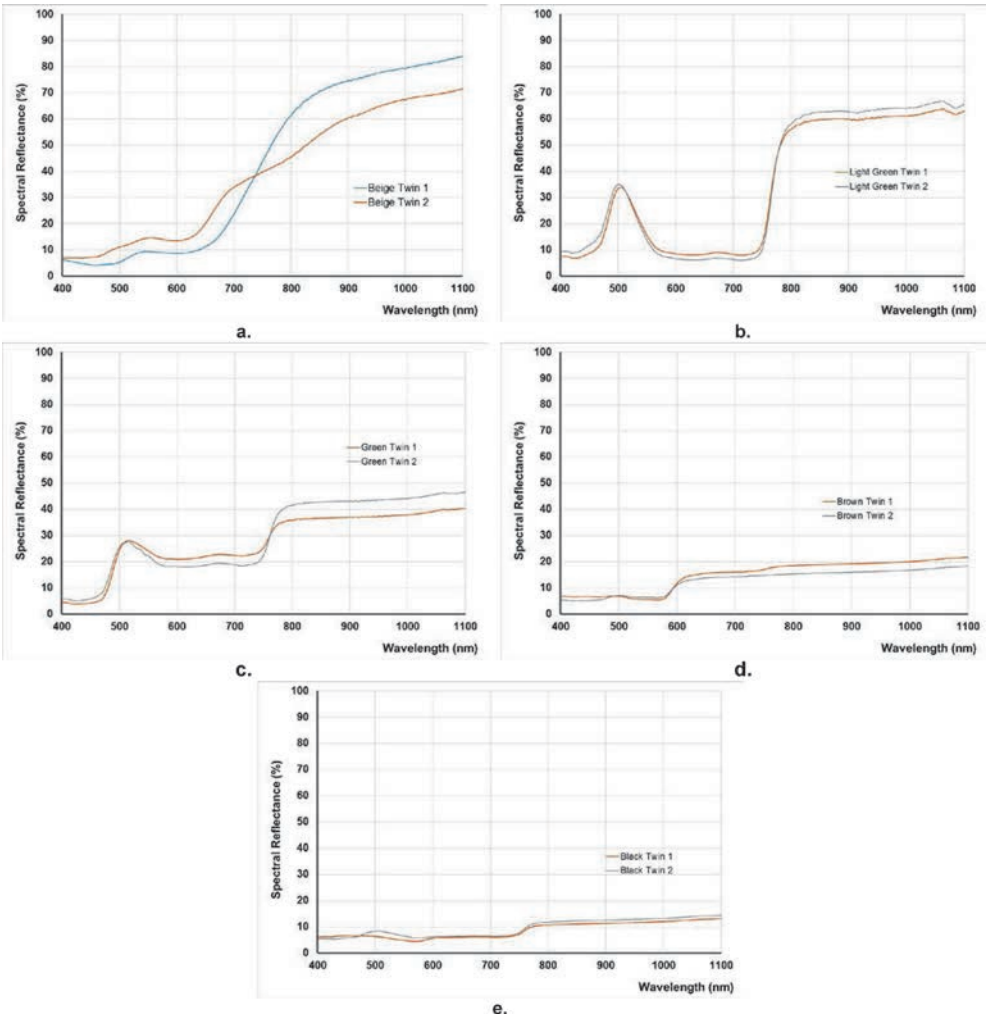


Fig. 9(a – e) – Remission plots of "Twin" dyes used in creating a pattern with hidden message – beige (a), light green (b), green (c), brown (d) and black (e), printed on textile substrate

The mixing of CMYK colours are calculated in order to assure the remission properties of "Twin" dyes as close as possible since it is impossible to achieve perfect match of curves. When the aimed colours are defined by mixing of process colours and printed on certain substrate, the spectrophotometric measurement is performed in order to determine the spectral characteristics (remission properties) of produced colours. Since the measurement and remission determination is always done after printing and from the print outs, the process of selection and determination of satisfactory "Twin" dyes is long term iterative process. Hidden information produced by Z_{40} dyes is surrounded, in pattern observed in VIS spectrum, by Z_0 dyes which have no response in NIR spectrum (at 1000 nm). In visual spectrum, Z_0 and Z_{40}

dyes must be equal in their spectral characteristics in order to assure invisibility of hidden „Twin" Z_{40} in visual experience of colours.

After analysis of the properties of colours produced and determination of satisfactory "Twin" dyes, the masking pattern with hidden graphic has been produced and printed on textile fabric using inkjet pigment printing technology (Hewlett Packard Designjet HP 5000). The pattern and the hidden message are shown on Figures 10a,b.

Figure 10: Military pattern with hidden message: 10a – image in VIS waveband (camouflage pattern); 10b – image in NIR waveband at 1000 nm (hidden message occurs, invisible to a naked eye in VIS waveband).

In visible spectrum the image is observed as camouflage pattern while in IR spectrum the used colours does not have response at 1000 nm and are not visible as a pattern. In the range of NIR waveband, only the hidden message is visible, which cannot be observed by the eye of an observer in visible spectrum. The testing by recording with "Z – RGB" camera (RGB camera, which records the RGB colours /red, green, blue/ in the visible light from 400 to 700 nm and Z camera which registers NIR and which measures the absorption of light in the Z size at 1000 nm), has been performed (Figure 10b). The official Croatian Coat of Arms included in pattern as individualized hidden message, is detected by recording with Z RGB camera in NIR spectrum at Z 1000 nm.

4. Conclusion

Acceptable colour scheme, which satisfies the requirement of camouflage and disruptive properties, is the one that reflects visible and NIR radiation in a way that mimics the reflectance spectrum of the natural background. The critical value in current camouflage concerns the reflection values of the colours in the Near Infrared (NIR) spectrum against NVG's (Night Vision Goggles) or NVD's (Night Vision Devices), as well as the SWIR (Short Wave Infrared) on the newer NVD's.

For an object to be successfully camouflaged it needs to closely match the environment's colours which is not an easy task regarding a multiple different scales of shades present in natural environment. Dyes applied to fabrics should reflects visible and NIR radiation in a way that mimics the reflectance spectrum of the natural background. With the development of IR detection systems it became necessary to take into account the IR reflectivity of paints and garments in the near IR region (700-1200 nm). Therefore, the camouflage patterns should be painted or printed with selected dyes and pigments having IR reflectivity that matched the expected surrounding topography. It has been determined that the flora does not absorb infrared light wavelengths, meaning that this part of a nature is "white" for IR detection devices. So, to achieve a masking effect, both in visual and NIR spectrum, the pattern should be created as multicoloured scheme of shades matching specific palette of nature, while in NIR spectrum the complete scheme must have no response and should appear white to IR detection equipment. Consequently concealing ability of textile camouflage in VIS and NIR spectral ranges mostly depend on the surface colour characteristics, particularly on the spectral reflectance distribution in these spectral ranges.

As far as the design process is concerned, there is a constant scientific study into the statistics of natural images, texture, visual perception and psychology, in order to achieve more and more effective camouflage patterns that should contain details at multiple spatial scales and should be similar in composition to natural images. For an object to be successfully camouflaged it needs to closely match the environment's colours, shapes and textures in order to reduce the detection probability. Since the presented patterns have been developed for woodland environment it was important to achieve effect that would mimic a specific colour relations and contrasts and the effect of dynamic colour relations, contrast and movements. It was confirmed that multiscale and disruptive pattern that consists of both micro and macro elements, with shapes broken into a smaller design elements with smoother geometry of boundaries between contrasting elements, will create more disperse and dynamic appearance closer to natural images and would be more effective when viewed against contrasting areas of real nature environment.

In this work, the example of the camouflage pattern is made through INFRARE-DESIGN® method. The progress was accomplished to achieve an effect of a camouflage uniform in the visual and also in infrared spectrum, since the detection of infrared communication is the given communication military standard. For the colours in this work, Twins Z_0 and Z_{40} were designed. In the visual spectrum Twins Z0 and Z40 should be equalized. Measurements are proposed for real print on textile. For each printing technique, for each material and colour, it is necessary to calculate the new colour settings and the new pairs of Twins. Targeted infrared marking on clothes and hidden individualized information are detected by infrared cameras and with especially designed ZRGB cameras for simultaneous visual and infrared

imaging at Z 1000 nm. This approach allows the installation of IR protection into the existing design, into the new design, into the design for a special action and as well the addition in the form of labels on a military uniform. A task of management with dyes, matter, simulation of the environment of the flora and camouflage for simultaneous viewing with ZRGB cameras, which are designed for this range, was achieved. The camouflage pattern is designed to be incorporated in the visual and infrared spectrum. Within the infrared design, hidden images, invisible to our eyes, as well the invisible information and labels are incorporated. A new way of identification, differentiation, labelling and hiding has been achieved and a method for proving the authenticity of camouflage clothing have been developed.

References

[1] Toet, A.; Hogervorst, M. A.: Urban Camouflage Assessment through Visual Search and Computational Saliency, *Optical Engineering*, **Vol**. 52 (2013) No. 4, pp. 041103-1 – 041103-8, ISSN: 0091-3286

[2] Rubeziene, V.; Padleckiene, I.; Baltusnikaite, J.; Varnaite, S.: Evaluation of Camouflage Effectiveness of Printed Fabrics in Visible and Near Infrared Radiation Spectral Ranges, *Journal of Materials Science*, **Vol. 14**, (2008) No. 4, pp. 361 – 365, ISSN: 0022-2461

[3] Rubeziene, V.; Minkuviene, G.; Baltusnikaite, J.; Padleckiene, I.: Development of Visible and Near Infrared Camouflage Textile Materials, *Journal of Materials Science*, **Vol. 15**, (2009) No. 2, pp. 173 – 177, ISSN: 1329-1320

[4] Baumbach, J.: Colour and Pattern Composition to Blend Objects into a Natural Environment, *The Proceedings of Colour – Effects and Affects, Interim Meeting of the International Colour Association*, Bergstörm, B. (Ed.), ISSN 0280-2198, Stockholm Sweeden, June 2008, Stockholm, (2008)

[5] Fakin, D.; Ojstršek, D.: The Effect of Camouflage Shades on Clothing Low Visibility, *The Proceedings of Colour – Effects and Affects, Interim Meeting of the International Colour Association*, Bergstörm, B. (Ed.), ISSN 0280-2198, Stockholm Sweeden, June 2008, Stockholm, (2008)

[6] Parac-Osterman, Đ.; Glogar, M. I.:The Characteristics of olive green shaded militarv clothes in nature surrouding//*Book of Proceedings of the 4th International Textile, Clothing and Design Conference*/Dubrovnik, 2008. pg. 991-996

[7] Glogar, M. I.; Žiljak Stanimirović, I.; Parac – Osterman, Đ.: The New Elements of Military Camouflage Design in VIS and NIR Spectrum, *Proceedings of XIIIth International Izmir Textile and Apparel Symposium*, IITAS, Kumbasar, E. Perrin Akcakoca ; Cay, Ahmet, et al. (Ed.)., ISBN: 978-605-338-043-6, Turkey, April 2014, Izmir, (2014)

[8] Glogar, M. I.; Žiljak Stanimirović, I.; Parac – Osterman, Đ.: Some Aspects of Smart Elements in Conteporary Military Uniforms, *Book of Proceedings of the 8th Central European Conference on Fiber – Grade Polymers, Chemical Fibers and Special Textiles*, Dekanić, T. and Tarbuk, A. (Eds.), ISBN 978-953-7105-63-1, Croatia, September 2015, Zagreb

[9] Marr, D.; Nishihara, H.: Visual information Processing: Artificial Intelligence and the Sensorium of Sights, *Technology Review*, **Vol. 81** (1978) No. 1, pp. 2-23

[10] Ziljak, V.; Pap, K.; Ziljak, I.: CMYKIR Security Graphics Separation in the Infrared Area, Infrared Physics and Technology, Vol. 52 (2009) No. 2-3, pp 62-69, ISSN: 1350-4495

[11] Ziljak, I.; Pap, K.; Ziljak Vujic, J.: Infrared Design on Textiles as Product Protection, Tekstil, Vol. 58 (2009) No. 6, pp 239-253, ISSN: 0492 – 5882

[12] Ziljak, V.; Pap, K.; Ziljak, I.: Infrared Hidden CMYK Graphics, The Imaging Science Journal, Vol. 58 (2010) No.1, pp 20-27, ISSN: 1368-2199

[13] Ziljak, V.; Pap, K.; Ziljak Stanimirovic, I.: Development of a Prototype for ZRGB Infraredesign Device, Technical Gazette, Vol. 18 (2011) No.2, pp 153-159, ISSN 1330-3651

[14] Pap, K.; Ziljak, I.; Ziljak Vujic, J.: Image Reproduction for near Infrared Spectrum and the Infraredesign Theory, Journal of Imaging Science and Technology, Vol. 54 (2010) No. 1, pp 1-9, ISSN 1062-3701

[15] Ziljak, V.; Pap, K.; Ziljak Stanimirovic, I.; Ziljk Vujic, J.: Managing dual colour properties with the Z – parameter in the visual and NIR spectrum, Infrared Physics &Technology, Vol. 55 (2012) No. 4, pp 326-336, ISSN: 1350-4495

[16] Boyd, R. J.: Colours of Northern Australia: Visible and Near – IR Reflectance of Natural Terrain Elements, *Available from* http://www.dsto.defence.gov.au/publications/scientific_record.php?prefix=1947&record=4225; *Accessed:* 2014-02-15

Scientific paper

The Improvement of Cotton Flame Retardancy with Aluminosilicate Application by Synergism

Ana Marija Grancarić*, Anita Tarbuk, Lea Botteri

University of Zagreb, Faculty of Textile Technology

The cellulosic textiles, as the most commonly used textile materials, have to be treated with those flame retardants (FR) that increase char or non-combustible products, usually organic phosphorous-based FR. Recently, the application of silicone based compounds has been investigated in order to lower/eliminate the use of toxic FR. Silicone-based compounds have an excellent thermal stability and high heat resistance with very limited release of toxic gases during the thermal decomposition. For this purpose, dyed cotton fabric has been treated with natural zeolite added to the bath with conventional FR. Natural zeolite (clinoptilolite) was applied in three different concentrations, whilst the conventional FR based on organophosphonate, and crosslinking agent based on melamine resin were applied in full and a half concentration. Such finished cotton fabrics were characterized by Fourier Transform Infrared – Attenuated Total Reflectance (FTIR-ATR) spectroscopy; thermal properties were determined by using Micro Combustion Calorimeter (MCC) and Thermogravimetric analysis (TGA); and burning behaviour vertically oriented specimens was determined by the bottom edge ignition according ASTM D6413-13, and the candle-like ignition according to ISO 4589:1996 in Limiting Oxygen Index (LOI) Chamber (Dynisco). Additionally, the fabric hand (subjective and objective evaluated), strength (breaking load, elongation), and aesthetic (colour difference) appearance were investigated. The aluminosilicates added to the bath contributed to flame retardancy of fabrics. Furthermore, treated cotton fabrics show high protective properties after 3 washing cycles. However, the high concentrations of activated zeolite affected other fabric properties.

Key words:

Cotton, flame retardancy, organophosphate, zeolite, LOI, TGA, MCC

*amgranca@ttf.hr

Introduction

Cotton fabric and blends is the most commonly used textile material in design and production of protective clothing, temporary and permanent drapes, bed covers, mattress covers, furniture fabrics, accounting for more than 60 % of the world annual consumption. However, as mainly consisting of cellulose, cotton textiles are the most flammable materials. In the case of fire, cotton materials represent a major risk as burning strong and fast. Therefore, it is necessary to improve its flame retardancy. According to Horrocks (2005) all flame retardant cottons are produced by chemically after-treating fabrics as a textile finishing process which, depending on chemical character and cost, yields flame retardant properties having varying degrees of durability to various laundering processes. Cellulosic textiles are treated with those flame retardants that increase char or non-combustible products, usually organic phosphorous-based flame retardants. Phosphorus-based compounds work in vapour or condensed phase by synergistically promoting the phosphorylation of cotton fabrics and enhancing the formation of protective char. Its flame resistance depends on the phosphorus concentration on the fabrics, and the presence of nitrogen if applied, as a result of nitrogen and phosphorus synergism (Baker & Drews 1985; Yoshioka-Tarver et al. 2012).

Recent papers investigate application of silicone based compounds as silicones, silicates and natural zeolite in order to eliminate the use of toxic flame retardants (Alongi et al. 2012, 2014; Grancarić et al. 2012, 2015). These compounds have an excellent thermal stability and high heat resistance with very limited release of toxic gases during the thermal decomposition. Since FR fabrics normally require a 20-40 wt% add-on of chemicals to perform efficiently, in this paper the application of organophosphorous-based flame retardant in full and a half concentration with and without addition of natural zeolite was researched for improving the flame retardancy of cotton fabrics.

Natural zeolite, e. g. clinoptilolite, is a microporous hydrated crystalline structure containing alumina-silicates with well-defined structures containing AlO_4 and SiO_4 tetrahedral linked through the common oxygen atoms. It has unique absorption and catalytic properties and for that reasons multiple uses in industry, agriculture, water purification and detergents. In medicine they are attributed antitumor, antiallergic, antiseptic, antireumatic and other properties. Tribomechanical activation of natural zeolite in patented machine results in zeolite grinding to submicron levels (micro- and nano-particles) resulting in increment of their reactive ability several hundreds of times. These activated natural zeolites are nontoxic substances, excellent for UV-R and microbes protection, for adsorption of proteins and small molecules such as glucose. It has strong anti-oxidative properties, acts as an immunoactivator and has decontamination effects. As an anti-oxidant it prevents forming of free radicals (FR) in human body. Therefore, clinoptilolite is used by cancer patients and pa-

tients with autoimmune diseases (Pavelić et al. 2002; Ivković et al. 2004). Applied externally in powder form, it quickens the healing of wounds and surgical incisions. Applied on textile material it increase active surface area, absorbs and scatters the UV-R, and gives off good antimicrobial and antifungal protection (Grancarić et al. 2007, 2009, 2012; Tarbuk et al. 2015). However, applied in high concentrations of activated zeolite also impart negative effects on fabric hand and colour. Besides of fabric hand, fabric drape is a property of fabrics associated with the aesthetic appearance of garments. Whereas high concentration of flame retardants imparts negative effects on fabric hand and drape, in this paper a critical assessment was performed after zeolite and conventional organophosphonate finishing.

2. Materials and Methods

Plain woven fabric of 100 % cotton, dyed in dark green shade, and mass per surface area of 165 g/m^2 was used. The fine powder of natural zeolite – clinoptilolite (micro and nanoparticles) from Slovakia was obtained by tribomechanical activation. Tribomechanically activated natural zeolite contained approximately 80 % of clinoptilolite and 20 % of silica, montmorillonite and mordenite. Detailed composition is listed in Table 1 (Pavelić et al. 2002).

Zeolite was applied in three concentrations – 1 g/l, 5 g/l and 10 g/l. It was applied by itself or added to the bath containing: Apyrol CEP (Bezema) – flame retardant agent based on organophosphonate (Fig.1); Apyrol MH (Bezema) – crosslinking melamine resin for durable finishes, phosphoric (V) acid – curing agent, and non-ionic surfactant Kemonecer NI (Kemo) – wetting agent.

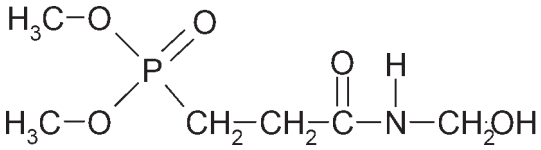


Fig. 1 – Dimethyl methylol phosphonamide acetic acid – Apyrol CEP (BEZEMA)

Fabrics were treated by pad-dry-cure procedure in Benz Pad-dry system. The cotton fabric was firstly padded (impregnated) in bath containing different flame retardant mixtures (Table 2), having wet pick up of 100 %. In continuous process impregnated fabric was dried at 110 °C for 2 min and afterwards cured at 155 °C for 4.5 min.

Table 1 – Composition and Physicochemical Properties of TMAZ analysed by ISEGA Forschungs- und Untersuchungsgesellschaft GmbH, Aschaffenburg, Germany (Pavelić et al. 2002)

Chemical composition	
Component	**%**
SiO_2	65,0-71,3
Al_2O_3	11,5-13,1
CaO	2,7-5,2
K_2O	2,2-3,4
Fe_2O_3	0,7-1,9
MgO	0,6-1,2
Na_2O	0,2-1,3
TiO_2	0,1-0,3
Si/Al	4,8-5,4
Empirical formula	
$(Ca,K_2,Na_2,Mg)_4Al_8Si_{40}O_{96} \times 24H_2O$	
Physicomechanical properties	
Specific mass	2,2-2,5 g/cm^2
Porosity	32-40 %
Effective pore diameter	0,4 nm
Ion-exchanging capacity	mol/kg
Total	1,2-1,5
Ca^{2+}	0,64-0,98
Mg^{2+}	0,06-0,19
K^+	0,22-0,45
Na^+	0,01-0,19
Ion-exchanging selectivity	
$Cs>NH_4^+>Pb^{2+}>K^+>Na^+>Mg^{2+}>Ba^{2+}>Cu^{2+}>Zn^{2+}$	
Chemicals absorbed	
NH_3, hydrocarbons $C_1 - C_4$, CO_2, H_2S, SO_2, NO_x, aldehydes	
Toxicity	
Nontoxic according to US Code of Federal Regulations (21 CFR 82, Subpart C)	

Table 2 – The labels and the composition of flame retardant mixtures in padding bath

Label	The composition of flame retardant mixtures
I	Unfinished – cotton fabric dyed in green shade
II	FR finish – 400 g/l Apyrol CEP; 70 g/l Apyrol MH, 24 g/l H_3PO_4, 1 g/l Kemonecer NI
III	FR finish – 200 g/l Apyrol CEP; 35 g/l Apyrol MH, 12 g/l H_3PO_4, 0.5 g/l Kemonecer NI
1, 5, 10	1 g/l, 5 g/l, 10 g/l of tribomechanic activated natural zeolite added to the bath
W	3 washing cycles

For durability fabrics were washed for 3 cycles at 40 °C for 30 min in Mathis according to ISO 6330:2000 – *Textiles – Domestic washing and drying procedures for textile testing* with ECE Test Detergent without phosphate.

The characterization of FR treated fabrics was performed by Fourier Transform Infrared – Attenuated Total Reflectance (FTIR-ATR) spectroscopy. 16 scans at a resolution of 4 cm^{-1} were recorded for each sample between 2000 cm^{-1} and 650 cm^{-1} on FTIR spectrometer (PerkinElmer, software Spectrum 100).

Since the flame retardant finishing has negative effect to the fabric properties, fabric mass per unit area, hand, strength and the color differences were determined according to standard methods.

Fabric mass per unit area, *m*, was determined according to ISO 3801:1977- *Textiles – Woven fabrics – Determination of mass per unit length and mass per unit area.*

Breaking force, *F,* and elongation,ε, were determined on TensoLab dynamometer acc. to ISO 13934-1:1999 *Textiles – Tensile properties of fabrics – Part 1: Determination of maximum force and elongation at maximum force using the strip method*.

A critical assessment of hand is usually relation between subjective assessment and objective measurement of eight physical properties – four related to bending, two to shearing and the other two, determined in this paper, coefficient of kinetic friction and drape coefficient, respectively. The subjective evaluation in this research was performed according AATCC Evaluation Procedure 5-2011. Fabric Hand: Guidelines for the Subjective Evaluation of. 10 evaluators have ranked fabrics according the physical attributes of hand.

The coefficient of friction was measured on FRICTORQ (Fig.2), fabric surface tester based developed and patented at the University of Minho, Portugal, by Lima *et al.* (2003). The procedure of dynamic or kinetic friction coefficient, μ_{kin}, measurement – a circular fabric sample of radius, *r,* is clamped and forced to rotate

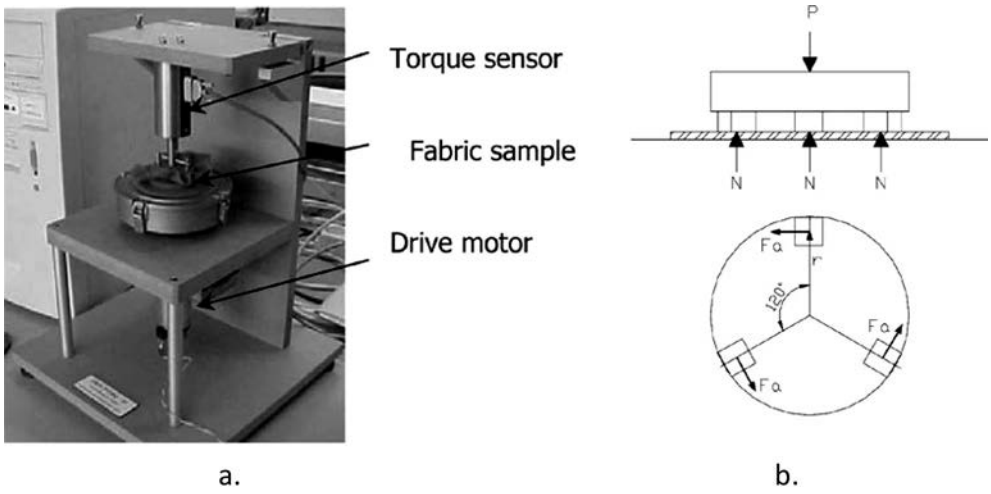


Fig. 2 – FRICTORQ II; a. the fabric surface tester, b. model (Lima et al. 2003)

around a vertical axis at a constant angular velocity while a vertical load, *P,* is concentrically applied by a static upper body by means of three small contact sensors, placed in a circle at 120°.

Friction coefficient is then proportional to the level of the dragging torque, *T,* measured on a high precision reaction torque transducer. Providing a relative displacement of approximately 90°, it is assured that a new portion of fabric is always moved under the sensors. Contact pressure of approximately 3,5 kPa is constant during the test. The torque signal is digitalized through an electronic interface and fed into a PC where dynamic or kinetic friction coefficient, μ_{kin}, is worked out and displayed, according to:

$$\mu_{kin} = \frac{T}{P \cdot r} \tag{1}$$

Fabric drape describes how the material falls or hangs over the 3-D form. It is the extent to which a fabric will deform when it is allowed to hang under its own weight. In this paper it was determined according to ISO 9073-9:2008 – *Textiles – Test methods for the nonwovens – Part 9: Determination of drapability including drape coefficient*, Method A. A circular specimen of the fabric being tested is held horizontally between smaller concentric discs and the exterior ring of fabric is allowed to drape into folds around the lower supporting disc. The shadow of draped fabric is cast from below onto ring of paper of known mass and of the same size as the unsupported part of the test fabric. The outline of the shadow is traced onto the paper ring, which is then cut along the trace. The mass of the inner part representing the shadow, m_{SA} and m_{PR} is mass of the paper ring [g] are determined. The drape coefficient was calculated acc. to:

$$D_{30} = \frac{m_{SA}}{m_{PR}} \cdot 100 \quad \textbf{(2)}$$

Remission spectrophotometer SF 600 PLUS CT (Datacolor) was used for measuring spectral values and colour attributes: H* (hue), L* (lightness) and C* (chroma). The difference in colour attributes (ΔH*, ΔL*, ΔC*) and the total colour difference (ΔE) were calculated automatically according to:

$$\Delta E^* = \left[(\Delta H^*)^2 + (\Delta L^*)^2 + (\Delta C^*)^2 \right]^{\frac{1}{2}} \quad (3)$$

Thermogravimetric Analysis (TGA) was performed on TGA Pyris1 (PerkinElmer). 5 mg of sample was stacked in an open platinum pan. The mass loss as a function of the temperature was measured in the air atmosphere within the range from 45 to 650 °C with changes of 30 °C/min.

Microscale Combustion Calorimeter (MCC) tests were done according to ASTM D 7309 on MCC-2, Govmark, USA. The 5 mg of test sample was placed within the sample cup and test is performed under following conditions: pyrolyser operating temperature range of min. 25 to 600 °C at a heating rate of 1 K/s in an inert gas stream (nitrogen, 80 ml/min); detection sensitivity limit of min. 5 mW, repeatability of ± 2 %.

Burning behaviour of fabrics was determined on vertically oriented specimens by the bottom edge ignition according ASTM D6413/D6413M-13b *Standard Test Method for Flame Resistance of Textiles (Vertical Test*), and by the candle-like ignition, characteristic for Limiting Oxygen Index (LOI) determination, according ISO 4589:1996 – *Plastics – Determination of burning behaviour by oxygen index* in LOI Chamber (Dynic).

Recently, it has been demonstrated that it is possible to quantify the synergism between two species when applied together as flame retardants in cotton finishing. The method proposed by Harrick's et al. (2005) is based on the calculation of synergism effectiveness (SE) parameter and can be calculated according to Grancarić et al. (2015):

$$SE = \frac{(Fp)_{fr+s} - (Fp)_p}{\left((Fp)_{fr} - (Fp)_p\right) + \left((Fp)_s - (Fp)_p\right)} \quad (4)$$

where (Fp) is a given flammability parameter (from flammability or combustion tests), $(Fp)_p$ is the flame-retardant property of the polymer alone (cotton), $(Fp)_{fr}$ is that of the polymer and flame retardant, $(Fp)_s$ is that of the polymer treated with the synergist, and $(Fp)_{fr+s}$ is that of the full formulation comprising flame retardant and synergist. In particular:

SE > 1	synergy is occurring;
0 < SE = 1	simple additive effect is occurring;
SE < 0	antagonism is occurring.

Therefore, in this paper the synergistic effect between organophosphate flame retardant and natural zeolite was researched and quantified based on the results of Limiting Oxygen Index.

3. Results and Discussion

The characterization of surface and chemical composition after FR treatment of cotton fabrics was performed by FTIR-ATR spectroscopy. From Fig.3 can be seen that all spectra have characteristic peaks for the availabilities of –OH groups (3300 cm^{-1}) and C–H bonds (2900 cm^{-1}). In the fingerprint area between 1450 cm^{-1} and 850 cm^{-1} the characteristic peaks are recorded at 1155 cm^{-1}, 1105 cm^{-1}, 1050 cm^{-1}, 1025 cm^{-1}, 1005 cm^{-1}, 985 cm^{-1} and 895 cm^{-1} for non-finished fabric (I-0). These peaks are characterized by: 1425 cm^{-1} bending CH_2 (sym) at C(6); 1155 cm^{-1} C–C asymmetric bonds; 1105 cm^{-1} C–O–C bridges, glycoside bonds; 1050 cm^{-1} C–OH secondary alcohols; 1025 cm^{-1} C–OH primary alcohols; 1005 cm^{-1} and 985 cm^{-1} –CH– bonds; and 895 cm^{-1} C(1)–O–C(4) symmetric bonds. Comparing the spectra shown in Fig. 1, it is evident that only organophosphate compound lead to significant changes in the functional groups of cotton. The peak at the 1637 cm^{-1}, corresponding to the absorbed water to -OH group of unfinished cotton fabric (I), is wider for cotton fabrics treated with organophosphate compound (II & III), indirectly indicating the increase of the number of -OH groups. The intensity of the band at 1204 cm^{-1}, corresponding to the COH in plane on the C(6) atom is reduced, indicating that this is precisely the atom occurred with reactive crosslinking agent. The new band at a wavenumber of 828 cm^{-1} belongs to the vibration P-CH_2 groups, which directly indicates that crosslinking has occurred with melamine and organophosphate compound.

The band between 1200 and 1250 cm^{-1} (1230 cm^{-1}) corresponds to the vibration bond P = O; the bands 1659 cm^{-1} to amide I (stretching vibration C=O bond) and 1540 cm^{-1} amide II (bending bond N-H). The P=O peak in the range of 1170-1230 cm^{-1} and P-O-alkyl peaks in the range of 1150-1200 cm^{-1} and 1330-1430 cm^{-1} are not so clear due to the high intensity C-O stretching, O-H bending, and C-H bending peaks from cellulose.

The results of mass per unit area determined acc. ISO 3801:1977 are shown in Fig.4.

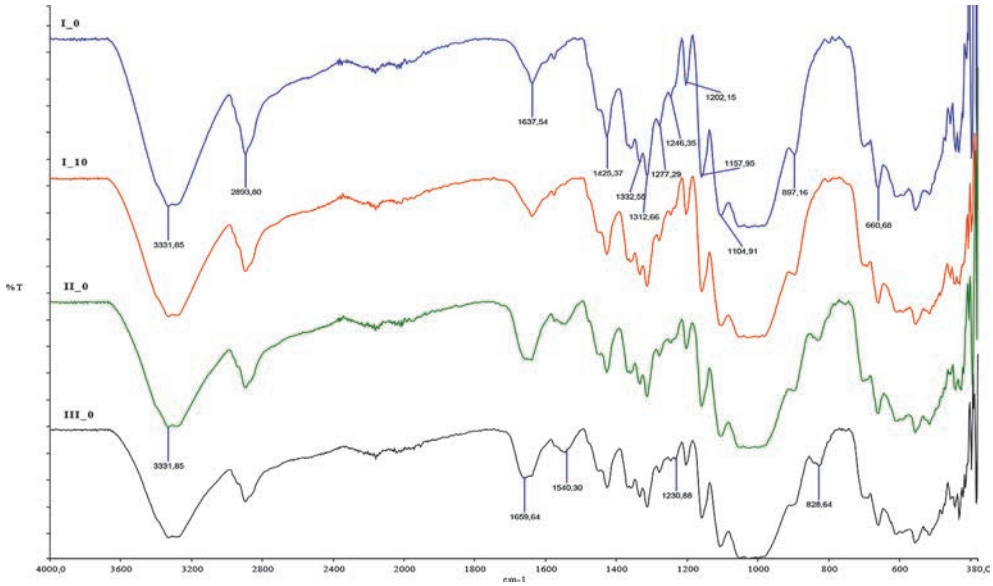


Fig. 3 – FTIR spectra of cotton fabrics: unfinished (I_0), treated with 10 g/l zeolite (I_10), and finished with organophosphate compound in full (II) and half concentration (III)

The results confirm that the flame retardants (FR) were bonded to the fabric surface. All finished fabric have significantly higher surface mass even after 3 washing cycles. The reduction in mass after washing was expected since the non-bonded

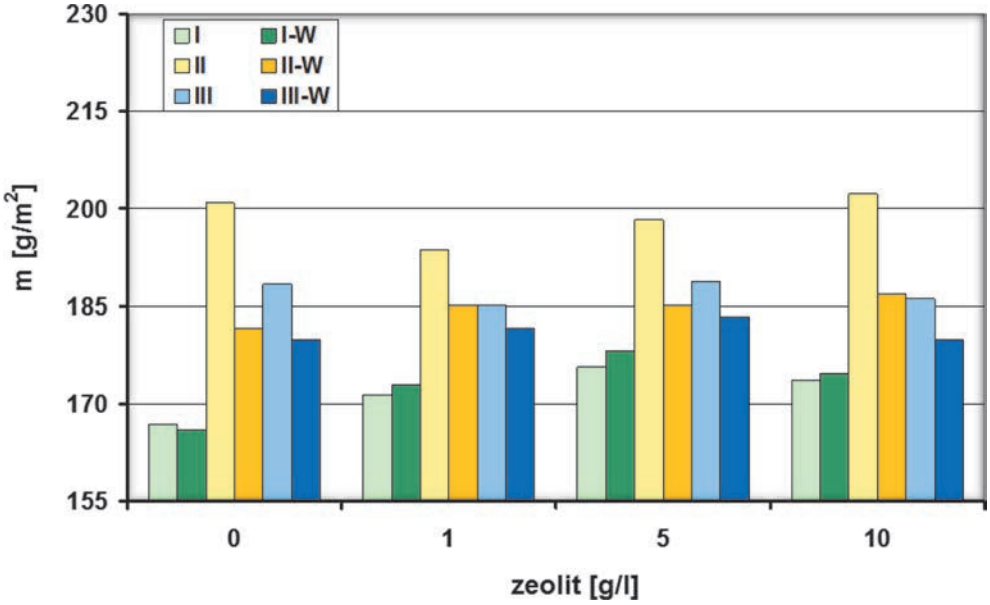


Fig. 4 – Mass per unit area of untreated, finished and washed cotton fabrics

FR was washed off the fabric. Treatment with organophosphate FR in full concentration (II), surface mass is the highest. Zeolite addition results in not so high mass increment, suggesting the inability of full spreading and bonding of organophosphate FR due to the reservation of places with adhesion bounded zeolite. The results of tensile properties are shown in Fig. 5.

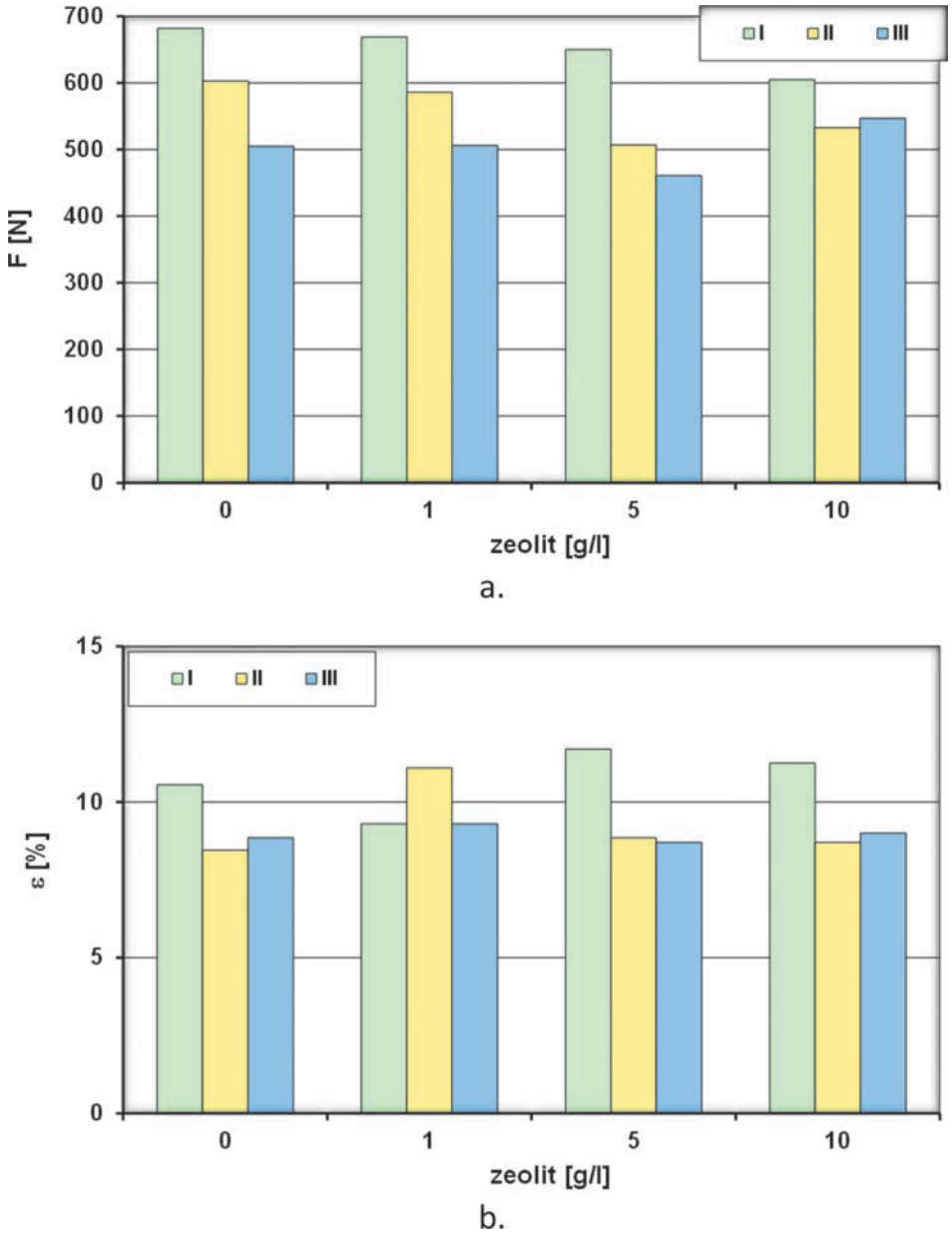


Fig. 5 – Tensile properties of finished cotton fabrics; a. breaking force, *F*, and b. breaking elongation, *ε*

It is well known that FR finishing lower the strength of fabric due to the crosslinking. From Fig. 5 can be seen that by zeolite application, irrespective of the amount, strength slightly decreases. It was expected due to the incorporating of zeolite nanoparticles between the cellulose chains with the consequence of the occasional tear of the hydrogen bonds.

As above said, fabric hand is associated with the aesthetic appearance and comfort of garments. Therefore, subjective evaluation was performed according to AATCC EP 5 guidelines; and objective by determining two physical properties – kinetic friction coefficient, m_{kin}, and drape coefficient, D_{30}. Results are presented in Table 3 and in Figure 6. From Table 3 it is quite evident the FR finishing affects fabric hand. All the evaluators have ranked fabrics from none (I_0) to the highest amount of zeolite (I_10) regarding all physical attributes of fabric hand.

Table 3 – Fabric hand subjective (Compression, Bending, Shearing, Surface Attributes) evaluation

Evaluation	Subjective			
	Compression	Bending	Shearing	Surface
Fabric	Attribute and rank			
I_0	soft	limp	firm	smooth
I_1	↓	↓	firm	↓
I_5	↓	↓	firm	↓
I_10	little bit hard	little bit stiff	firm	rough
II_0	soft	little bit stiff	firm	smooth
II_1	↓	↓	firm	↓
II_5	little bit hard	↓	firm	↓
II_10	hard	stiff	firm	rough
III_0	soft	limp	firm	smooth
III_1	↓	↓	firm	↓
III_5	↓	↓	firm	↓
III_10	little bit hard	little bit stiff	firm	rough

Applying full concentration of FR Apyrol CEP (II) and a small amount of zeolite hand is worse than when applied half the concentration. Same result was obtained by objective measurement of fabric drape and friction. Zeolites, as micro- and nano-particles, penetrate into the cotton fabric, but most of those particles remain on

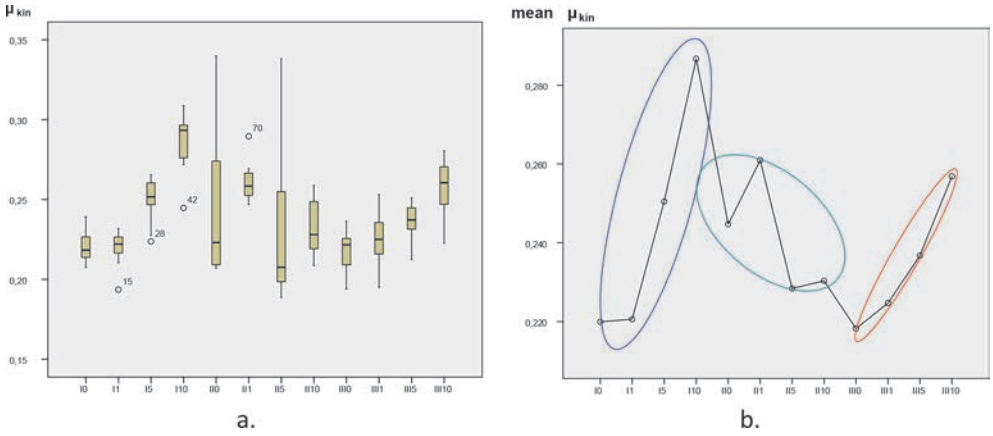


Fig. 6 – Dynamic friction coefficient, μ_{kin}, of finished cotton fabrics; a. Box-plot and b. mean values

the fabric surface. Therefore, friction is higher and dynamic friction coefficient, μ_{kin}, increases with zeolite concentration. Applying small concentration of zeolite (1 g/l) there is no significant increment of friction coefficient. Raising the concentration of zeolite, μ_{kin} mean values rapidly increase. The box-plot of Figure 6 for the samples summarizes the observations previously discussed: Samples I and III show a similar trend of an increased friction coefficient with increased zeolite concentrations. The experiments also exhibit a small dispersion of results. However, for samples II the trend is not clear and the dispersion of results is big, especially for samples II-0 and II-5. A more detailed experimental work might be recommended in order to better understand the influences of the various agents.

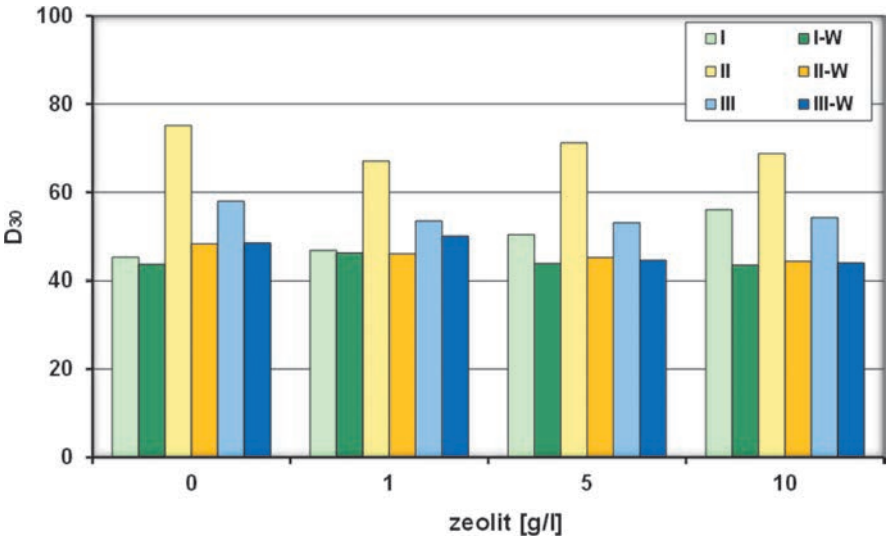


Fig. 7 – Drape coefficient of finished cotton fabrics

By measuring drape coefficient of zeolite and FR finished fabrics it is evident that the fabric is stiffer, and fabric flexibility reduced. This effect is enhanced for full concentration of organophosphate FR. Using zeolite flexibility does not significantly change. By washing the fabric part of un-bonded zeolite wash off the fabric, reducing surface mass per area and improve the hand. However, this affects the fabric protective properties.

The FR finish changes the fabric colour. The total colour difference (ΔE) was calculated automatically from differences in colour attributes. The differences are shown in Figure 8.

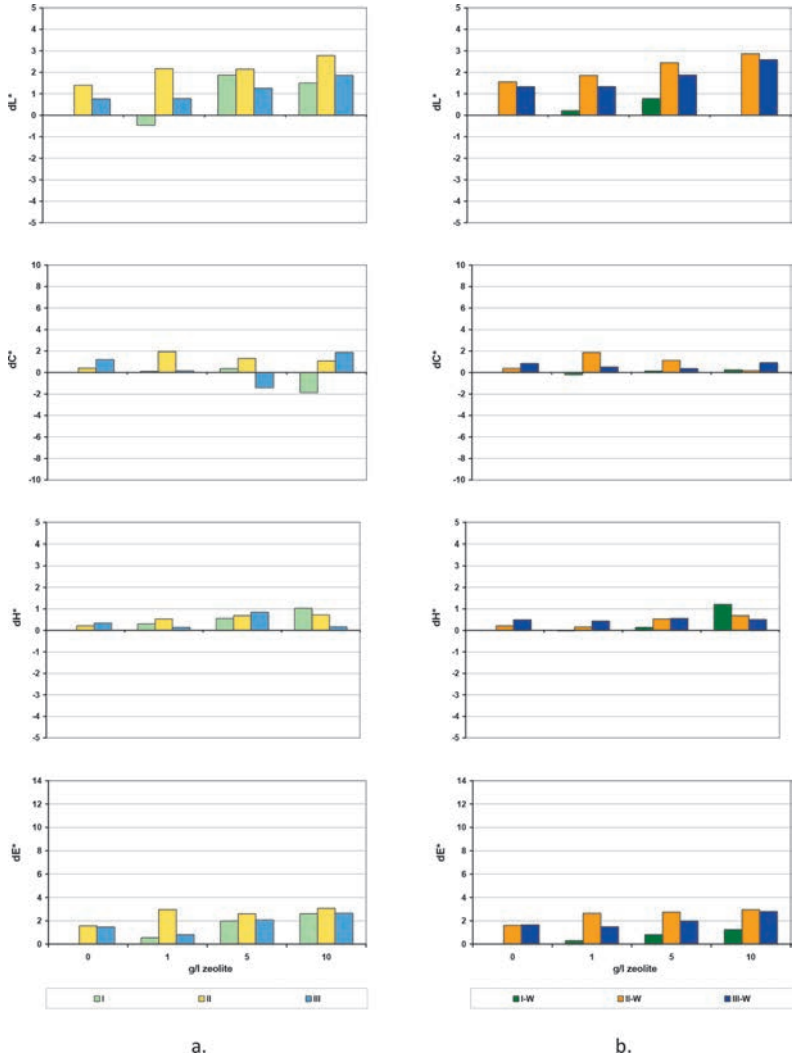


Fig. 8 – The colour difference of a. finished, b. washed cotton fabrics

Flame retardant and zeolite finishing change the colour of fabric. From Figure 8 it is evident that the zeolite addition, which is slightly yellow, leads to higher fabric lightness. The flame retardant application has even more enhanced change in colour, as a result of higher lightness. The washing process changes basic hue whilst some of not bonded agents is washed off, resulting in lower colour differences. All the values are acceptable except from finishing in full concentration (II) which results in $\Delta E^* \geq 2$, what is not acceptable in textiles.

Burning behaviour was determined according the standards after the fabric treatment and washing. According to the results of vertical test and Limiting Oxygen Index presented in Tables 4 and 5, it can be seen that unfinished cotton burns rapidly when present only 18 % of O_2. The conventional flame retardant for cellulose – organophosphate showed expected very high LOI value: LOI=33 for full and LOI=30 for half concentration. Application of zeolite has a positive effect on the flame retardancy of cotton. It improves LOI to 19, in open flame fabrics burn much slower and the afterglow is lower. However, treatment with conventional flame retardants with addition of zeolite significantly increased the LOI regardless of applied in full or reduced concentration. Considering the applied concentration, from the results of the length of char and LOI, it is evident that the concentration of 5 g/l is optimal one, since it results in the same, and in some cases, even better results than if applied 10 g/l of zeolite. Applying zeolite in the bath containing a

Table 4 – Burning behaviour of vertically oriented fabric according to ASTM D6413 / D6413M-13b (Vertical test) of treated and washed cotton fabrics

Fabric	after treatment			after washing		
	$t_{after\ flame}$ [s]	$t_{afterglow}$ [s]	l_{char} [mm]	$t_{afterflame}$ [s]	$t_{afterglow}$ [s]	l_{char} [mm]
I_0	24	22	300	15	13	300
I_1	24	9	300	15	7	300
I_5	28	8	300	11	7	300
I_10	23	7	300	13	12	300
II_0	0	0	115	0	0	180
II_1	0	0	105	0	0	180
II_5	0	0	75	0	0	145
II_10	0	0	115	0	0	175
III_0	0	0	120	0	0	175
III_1	0	0	110	0	0	180
III_5	0	0	110	0	0	160
III_10	0	0	125	0	0	185

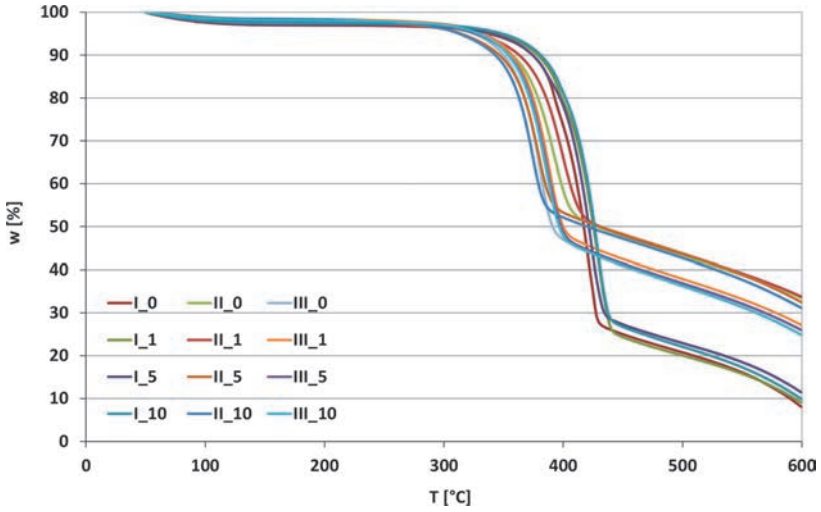
Table 5 – Burning behaviour and Limiting Oxygen Index, LOI according to ISO 4589:1996 of treated and washed cotton fabrics

Fabric	after treatment		after washing	
	$t_{100\ mm\ mark}$ [s]	LOI [%]	$t_{100\ mm\ mark}$ [s]	LOI [%]
I_0	64	18	64	18
I_1	117	19	130	18
I_5	120	19	99	19
I_10	124	19	102	19
II_0	25	33	19	33
II_1	31	33	15	33
II_5	21	34	42	33
II_10	28	33	25	33
III_0	28	30	23	30
III_1	25	32	22	31
III_5	30	34	25	32
III_10	25	33	18	32

half amount of flame retardant finish, results in almost the same flame retardancy as if applied full amount. Since FR fabrics normally require a 20-40 wt% add-on of chemicals to perform efficiently, application of zeolite to conventional FR makes this process ecologically and economically favourable. It is to point out that the protective properties are retained after 3 washing cycles.

For better understanding of the change in cotton thermal properties under the heat Thermo-gravimetric analysis (TGA) and Micro combustion calorimetry (MCC) were done. The results are shown in Figures 9-10 and in Tables 6-7.

From TGA curve shown in Figure 3 can be seen that cellulose has two-step mass loss. The first stage occurs at temperature above 110 °C and it is connected with evaporation of easily volatile compounds, water, impurity etc. At temperature above 110 °C, chemical bonds begin to break. Between 100 °C and 200 °C, non-combustion products, such as carbon dioxide, trace of organic compounds and water vapor, are produced. Above 200°C the cellulose starts to degrade, when the cellulose structure has absorbed enough energy to active the cleavage of the glycosidic linkage. The thermal degradation of cellulose takes place at 300–400 °C through two competitive processes, namely depolymerization and dehydration. Depolymerization is initiated by the scission of acetal bonds between the chain glycosidic units, followed by successive splitting of volatile levoglucosan, the cyclic



Fig″. 9 – TGA curves of FR treated cotton fabrics recorded in air, heating rate 30 °C/min

monomer of cellulose, from ensuing chain ends. It is known that other covalently-bonded organic molecules, such as reactive dyes and monochloroacetic acid, can reduce cotton flammability by interfering with the pyrolysis of cellulose to levoglucosan [2]. Since the used cotton fabric (I_0) was mill dyed with reactive dyestuff in dark green shade, its degradation takes place at higher temperature than 400 °C, but the degradation curve is the same.

Table 6 – Thermogravimetric data of untreated and FR treated cotton in the air

Fabric	$T_{onset\ 5\%}$ [°C]	T_{max1} [°C]	Residue at T_{max} [%]	T_{max2} [°C]	Residue at 650 °C [%]
I_0	329	410	45.3	585	8.1
I_1	335	418	43.5	584	9.0
I_5	335	411	48.9	584	11.4
I_10	351	417	45.6	576	9.9
II_0	320	382	66.6	579	34.5
II_1	321	388	67.5	582	33.8
II_5	304	369	67.6	571	32.5
II_10	306	365	66.4	556	31.7
III_0	318	370	64.6	540	26.1
III_1	323	378	63.2	539	27.1
III_5	320	375	65.3	559	26.0
III_10	321	375	64.1	583	24.9

Compete dehydration reactions lead to thermally stable aliphatic structures (char I), which subsequently are converted into aromatic structures (char II), with water, methane, carbon mono and dioxide evolution (450–600 °C). As temperature increase to around 600 °C, the production of volatile compound is completed. The continuing weight loss is due to degradation of the remaining char. The residue that remains (around 10 %) is an activated char that can be oxidized to carbon dioxide, carbon monoxide and water vapor. Cotton treated with conventional flame retardant organophosphonate (II and III) started the degradation at lower temperature (as demonstrated by the $T_{onset5\%}$ linear decrease, Table 6) and the maximum temperature of first weight loss (T_{max1}) is strongly reduced. The reason of that is the presence of phosphor and nitrogen groups. Further heat increment leads to inhibition of volatile species and promote char formation. When the concentration of organophosphonate was reduced to half (III_0) degradation started at higher temperature than if applied in full concentration. It can be assumed that in the first step III_0 produce more volatile species than II_0, but less than unfinished cotton (I_0) which was expected. In second step, when the oxidation starts, both samples were inhibiting volatiles and produce similar char formation.

From Figure 10 and Table 7 it is clear that cotton material treated with conventional flame retardant has much lower Heat Release Rate (HRR) and Specific Heat release (h_c). This confirms the lower total system energy and better thermal properties that

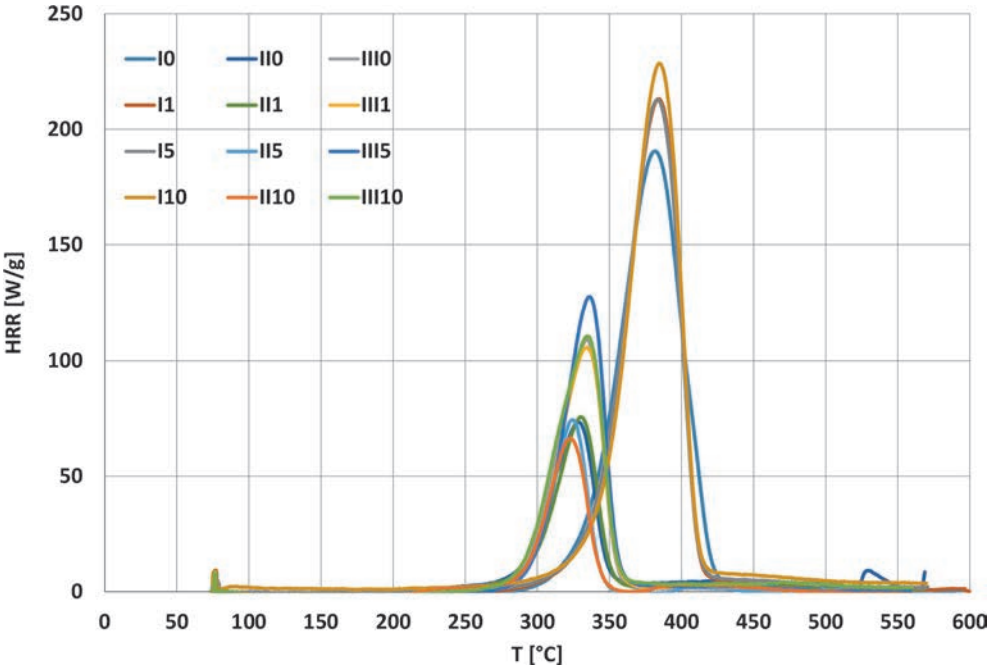


Fig. 10 – MCC curves of untreated and FR treated cotton fabrics

are crucial for flame retardancy in general. The Heat Release Rate (HRR) has slowly decreased by the addition of zeolite to the each bath. Combustion temperature of the treated cotton is lower and time to ignition is faster than of unfinished cotton. The higher the char yield, the more carbon/inorganic material was left behind, the decreased amount of combustible volatile release, resulting in lower flammability (lower H.R.R. temperature and H.R.R. values). From the MCC results it can be noticed that the samples that have lower Heat Release also produce less volatiles species and more char (Alongi et al. 2013). The MCC results confirm TGA results.

Table 7 – MCC data of untreated and FR treated cotton in the air

Fabric	Heat release capacity η_c (J/g-K)	Maximum specific heat release Q_{max} (W/g)	Heat release temperature T_{max} (°C)	Specific Heat release h_c (kJ/g)	Yield of pyrolysis residue Y_p (g/g)	Specific heat of combustion of fuel gas $h_{c,gas}$ (kJ/g)
I_0	204	207.24	388.41	10.4	0.06	11.08
I_1	209	212.98	362.53	10.2	0.09	10.58
I_5	209	212.25	386.75	10.0	0.11	11.22
I_10	221	224.50	387.90	10.2	0.11	11.42
II_0	71	72.54	331.15	3.05	0.31	4.42
II_1	72.5	74.42	333.45	2.75	0.34	4.18
II_5	71	72.34	266.58	2.9	0.33	4.19
II_10	62	63.11	280.47	2.9	0.30	4.43
III_0	105	106.73	287.57	4.9	0.23	6.08
III_1	101	102.44	291.51	4.7	0.22	6.14
III_5	121	122.67	293.47	4.6	0.23	6.16
III_10	108	109.99	289.89	4.6	0.19	6.02

In this paper the synergistic effect between FR organophosphate and natural zeolite clinoptilolite was researched and quantified based on the results of Limiting Oxygen Index. Synergistic effect, SE was calculated according to (4). Results are collected in Table 8.

From Table 8 it is evident synergistic effect between half concentration of Apyrol and zeolite as synergist (SE-III1=1,08, SE-III_5=1,23, SE-III_10=1,15). The synergistic effect remained after 3 washing cycles (SE-III_1=1,08, SE-III_5=1,07, SE-III_10_1,07).

When zeolite added to full concentration of Apyrol CEP only additive effect occurred.

Table 8 – Synergistic effect of organophosphate and zeolite

Fabric	Synergistic effect, SE after treatment	Synergistic effect, SE after washing treatment
SE-II_1	0,94	1,00
SE-II_5	1,00	0,94
SE-II_10	0,94	0,94
SE-III_1	1,08	1,08
SE-III_5	1,23	1,07
SE-III_10	1,15	1,07

4. Conclusion

Natural zeolite improves the cotton flame retardancy as fabrics burn much slower and the afterglow is lower. Added to the bath with conventional flame retardant zeolite significantly increase the LOI value. However, it should be point out that the highest concentration of applied zeolite (10 g/l) does not show any improvement concerning length of char or Limiting Oxygen Index as zeolite concentration of 5 g/l proved to be the optimal one.

Traditional high concentrations of organophosphonate retardants impart negative effects on fabric handle and drape. In the case of applying 50 % these concentrations in the flame retardant finishing, fabric handle and drape were considerably improved. However, achieved flame retardancy is almost the same as applied in the full concentration. In this case synergistic effect occurred what was quantified based on the results of Limiting Oxygen Index.

Therefore, it is to suggest using the lower flame retardant concentration with zeolite addition. It is important that the flame retardancy with these treatments is retained after three washing cycles.

Acknowledgements

The paper is a part of the research under the *COST Action project MP1105: FLARETEX*. The authors would like to thank the European Commission for the funding of the project.

References

[1] Alongi J. *et al*. 2014. Current emerging techniques to impart flame retardancy to fabrics: An overview. Polym. Degrad. Stab. Vol. 106 (8), 2014, pp. 138–149.

[2] Alongi, J., Camino, G., Malucelli, G. (2013). Heating rate effect on char yield from cotton, poly(ethylene terephthalate) and blend fabrics, Carbohydrate Polymers Vol. 92 (2) 2013, pp. 1327-1334.

[3] Alongi, J., Ciobanu, M., Malucelli, G. (2012). Sol–gel treatments on cotton fabrics for improving thermal and flame stability: Effect of the structure of the alkoxysilane precursor, Carbohydrate Polymers, Vol. 87 (1) 2012, pp. 627–663.

[4] Barker, R. H. & Drews, M. J. (1985). Flame Retardants for *Cellulosic* Materials, Chapter 17 in Cellulose Chemistry and its Applications, (Ed. T. P. Nevell and S. H. Zeronian), Ellis Horwood Limited, ISBN 0-85312-463-9, West Sussex, England, pp. 423-454.

[5] Grancarić, A. M. *et al.* (2012). Activated Natural Zeolites on Textiles: Protection from Radioactive Contamination, Chapter 8 in Intelligent Textiles and Clothing for Ballistic and NBC Protection Intelligent Textiles and Clothing for Ballistic and NBC Protection, Springer, ISSN 0353-9164, Heidelberg, pp. 157-176.

[6] Grancarić, A. M. *et al*. (2015). Synergistic effects occurring between water-glasses and urea/ammonium dihydrogenphosphate pair for enhancing the flame retardancy of cotton. Cellulose. Vol. 22 (4) 2015, pp. 2825-2835.

[7] Grancarić, A.M., A. Tarbuk, I. Kovaček, (2009). Nanoparticles of Activated Natural Zeolite on Textiles for Protection and Therapy, Chem. Ind. & Chem. Engineering Quarterly. Vol. 15 (4) 2009, pp. 203-210.

[8] Grancarić, A.M., Markovic L., Tarbuk A. (2007). Active Multifunctional Cotton Treated with Zeolite Nanoparticles. Tekstil Vol. 56 (9) 2007, pp. 543-553.

[9] Horrocks, A. R. (2005). Thermal (heat and fire) protection, In *Textiles for Protection*, pp. 398-440 (2005), Woodhead Publ. Ltd, Cambridge, England.

[10] Horrocks, A.R. *et al*. (2010). Quantification of zinc hydroxystannate and stannate synergiesin halogen-containing flame-retardant polymeric formulations. J Fire Sci, Vol. 28 (9) 2009, pp. 217–248.

[11] Ivkovic S, et al. (2004). Dietary Supplementation with the Tribomechanically Activated Zeolite Clinoptilolite in Immunodeficiency. Advances in Therapy Vol. 21, 2004, pp. 135-147.

[12] Lima M., L. Hes, J. Martins**,** *Proceedings of 3rd International Conference on Advanced Engineering Design*, 1-4 June 2003, Prague, Czech Republic.

[13] Pavelic, K. *et al.* (2002). Immunostimulatory effect of natural clinoptilolite as a possible mechanism of its antimetastatic ability. J Cancer Research and Clinical Oncology Vol. 128 (1) 2002, pp. 37-44.

[14] Tarbuk, A., A. M. Grancarić, S. Magaš (2015). Modified Cotton Socks – Possibility to Protect from Diabetic Foot Infection. Collegium Antropologicum Vol. 39 (1) 2015, pp. 177-183.

[15] Yoshioka-Tarver, M. *et al.* (2012). Influence of N-P Base Fiber Reactive Organophosphate Flame Retardant on Cotton Thermal Behavior, AATCC Review Vol. 12 (5) 2012, pp. 52-59.

Review paper

Flax – Does it Have a Future?

Maja Andrassy[1], Ružica Brunšek[1,*] and Jasminka Butorac[2]

[1]Faculty of textile technology, University of Zagreb,
Prilaz baruna Filipovića 28a, Zagreb, Croatia
[2]Faculty of Agriculture, University of Zagreb,
Svetošimunska cesta 25, Zagreb, Croatia

Flax belongs to the oldest cultivated plants in the world. For thousands of years it has been used for obtaining fibres, food and medicaments with numerous ups and downs. The principal areas in the world where textile flax has traditionally been cultivated, and to a large extent still is, are northwest and eastern Europe, China and Egypt. However, flax was used to produce textile products, which in the West had been progressively replaced by articles made from cheaper fibres such as cotton, jute and common man made fibres. It experienced a renaissance near the end of the 20th century, when ecological awareness and responsibility for the environment increased through the enhanced protection of natural resources.

Nowadays, many researchers all over the world, and also in Croatia, are trying to improve the cultivation of flax and to expand its usage. As flax fibre has not been grown in Croatia for more than 20 years, the investigations of flax is very scarce. Since 2001 Zagreb County, Ministry of Science, Education and Sports and Ministry of Agriculture, Fisheries and Rural Development have joined in the projects of conserving this crop and the cultural heritage of the Republic of Croatia. Therefore, it is necessary to introduce newly created agro technical and high-yield cultivars of flax fibre and to investigate their important agronomic and textile-technological properties under Croatian environmental conditions, with respect to the character and physiognomy of the existing plant production in particular areas of Croatia, but also to the socio-economic circumstances in those areas. This is also a precondition for the revival of flax fibre on Croatian family farms. It would produce a basic raw material for new domestic market products as an autochthon Croatian product. Certainly, the possibility of increased employment is no less important, bearing in mind flax production and processing.

[*]ruzica.brunsek@ttf.hr

Key words:

flax, flax fibres, renewal of flax production, cultivars, agronomic and textile properties

1. Introduction

Flax is one of the oldest textile fibres used by humankind. It was already known as a textile fibre in the Stone Age, while its seeds were used for food. Archaeological excavations of Stone Age lake-side dwellings of people in Switzerland, dated at approximately 7000 B.C., found flax seeds, twines and fishing nets (Šurina et al., 2009).

Flax was cultivated in Egypt already more than 6 000 years ago. Old Egyptian documents describe flax harvest and it's processing. Famous mummies were wrapped in flax fabric which has been preserved until today. The "Linen book of Zagreb" (Liber linteus Zagrebiensis), a manuscript with the longest preserved text in Etruscan language is the only remaining Etruscan text with literary characteristics and the only preserved sample of a linen book from the old ages. The canvas with about 1130 words has been preserved in five subsequent strips (Figure 1) (Preasted, 1906).

Fig. 1 – Mummy from Zagreb Archaeological Museum (Jeny, 2015)

Until the end of the 18th century flax fibres were, among wool and silk, one of the most important textile raw materials (Harris, 2004).

A rapid introduction of cotton in the 19th century and later the invention and development of man-made fibres in the first half of the 20th century, especially after the 2nd World War, were the reasons for considerable reduction of flax cultivation (Pasković, 1965). The attitude to this ancient culture unjustly neglected over the years changed towards the end of the 20th century in line with the general approach of greater responsibility towards environment and the need to protect natural resources and to support a more intensive use of relatively quickly renewable raw materials for the use in different industries (Franck, 2005).

Through past centuries flax fibre was traditionally grown and used in many regions of Croatia. Fibre flax culture has been practically abandoned in this country despite the increasing number of family farms, textile and other industries interested in the resumption of its production (Andrassy et al., 2004). According to the statistics at the beginning of 20th century there were approximately 9 000 hectares planted with flax in Croatia. Flax yields per hectares as well as flax fibres yield were very low at that time (Pasković, 1954) and the reasons can be found in facts that the greatest flax fibre manufacturers were family farms with deficient cultivation plans and insufficiently applied agro technical measures (fertilization, cultivating land, sowing, foliar fertilization and disease protection). Withought powerful flax fibre processing industry (only one flax processing mill at Cesko Selo, eastern part of Croatia) and possibility of manufacturing higher-quality flax, whose fibres could produce a better yarn for higher-quality flax fabrics, as well as neglect of traditional costumes caused by the introduction of cheaper cotton fabric and latter man-made fibres, caused flax culture and flax fibre production to be abandoned.

Since fibre production in Croatia (PES – Vartilen; wool at Koteks) were ruined during war in the nineties a need for the reviving of fibre production arises as strategic matter. Having in mind our tradition, new trends of fashion and composite industry as well as a fact that Croatia has possibilities for flax production an idea for renewal of flax production in accordance with sustainable development was born few years ago. This paper gives short overview on that matter.

2. Production of flax fibres

Compared to other textile fibres flax cultivation and manufacture requires less fertiliser and weed control chemicals than cotton. It is good rotation crop. It selectively absorbs heavy metal pollutants from contaminated soils. It grows in temperate cli-

mates where there is an abundance of good quality, available land (unlike cotton). It is biodegradable (unlike synthetic fibres). It requires no greater energy input during manufacture than do other fibres, and less than is required for synthetic fibres. Fibre flax is non-allergenic. It is comfortable to wear due to its rapid absorption and desorption of moisture. It is easily washed (Šurina et al., 2009; Butorac et al., 2006a).

Flax does not thrive as a monoculture, the plant uses soil nutrients, the occurrence of disease, pests and weeds is greater, and the accumulation of soil toxins is also greater. Flax can be planted on the same area every six to seven years. The quality of soil preparation determines sowing quality and the quality of flax harvesting (Butorac et al., 2006b; Butorac et al., 2009a).

Nitrogen fertilization is one of the most important agrotechnical interventions for growing fibre flax. Nitrogen is the most important plant nutrient that influences the development of plants and thus the developmental stages of the plant as well as the yield and fibre quality of flax fibre. Fibre flax requires less fertilising. Excess fertilizing and especially of nitrogen can encourage „lodging" of the crop. The quantities required are lower than for other fibre crops. To obtain an optimum yield of the flax steam of 8 to 9 t/ha corresponding to a fibre yield of 2 t/ha of flax, it should be fertilized with 35 kg/ha nitrogen, 100 kg/ha phosphorus and 150 kg/ha potassium (Šurina et al., 2012a; Butorac et al., 2014) . It is advisable not to sow fibre flax for the first time in fields where the level of mineral nitrogen is too high or where after recent manuring, the quality of undecomposed organic matter is too high. With regard to the expected yield, the spring nitrification capacity of the soil should be analysed and duly adjusted to the needs of the crop. It is unnecessary to supply any additional nitrogen, if the mineral nitrogen (at sowing time) and the additional spring nitrification produce a level, which is higher than the needs of the plant (Butorac et al., 2010a; Butorac et al., 2010b; Andrassy et al., 2010). The control of nitrogen levels is very important, as excess nitrogen not only increases the risk of lodging, but also decreases the yield and quality of fibres. If the fertilizer is added just before sowing, it can be an nitrogen, phosphorus, and potassium (NPK) compound where the form of nitrogen is half-nitric and half-ammoniac (Butorac et al., 2009a; Sharma and Sumere, 1992).

Phosphorus positively affects the development of the root and fibre strength and takes part in seed formation, while potassium stimulates fibre and seed formation. It creates a stronger bond among fibre bundles, and the stem is easier to process technically (Butorac et al., 2009a).

Effect of temperature. Fibre flax is a plant growing in wet and moderately warm regions. With respect to average annual air temperature, flax requires colder areas with moderate climate. The minimal temperature for seed germination ranges from 2 to 5 °C, and the optimal temperature ranges from 16 to 17 °C. Flax is especially sensitive to low temperatures at the beginning of emergence (Butorac et al., 2009b).

Effect of water. Fibre flax consumes a considerable amount of water during the vegetation period. Since flax grows in wetter and more moderate climates, in Croatia it could be grown in mountain regions and in lowland regions. These regions are abundant in moisture during their vegetation (> 600 mm precipitation per year). After sprouting flax does not require much water, but later water requirements become greater and greater, and it increases until flowering and decreases until maturity. The quantity of precipitation in May is crucial for fibre yield and quality in Croatian ecological conditions (Butorac et al., 2009b).

Effect of soil. Structural soils are suitable for flax cultivation, with good water and air regime (sandy loam, loamy sand soils which are porous and have a weak acid to alkaline reaction, pH 6.2-7.2). Due to lodging soils with high natural fertility, lime-rich soils, dry sandy soils, heavy clay soils with surface water and soils with a high level of underground water (Butorac et al., 2009b).

Sowing and harvesting time. The end of March or the beginning of April is the most favourable sowing time in the northwest of Croatia. In order to achieve the optimum assembly of 1 800 plants/m^2 in harvesting, in sowing about 2 500 fibre flax seeds/m^2 should be planted, depending on germination and seed purity. Flax is to be planted in weed-free soil. Vegetation period of fibre flax ranges from 100 to 120 days. The long strong bast fibres (bundles) for further processing are found in the stem, between the outer bark and a woody core. The root system does not contain fibre so it is not necessary to harvest the entire plant. The bundles lie around the core and are attached to it and one another by pectin's. The individual fibres, 10 to 40, are held together in the bundles by pectin's. The time of harvest is influenced by climatic conditions and the crop's final use. Harvesting of flax at different maturities provides a diversity of products: early crop for fibre only, with attached immature seeds for greater fibre quality, or mature crop for both seed and fibre production. Actually, there are three degrees in the ripening of the flax grown to make linen: green, yellow and brown (Butorac et al., 2012; Šurina et al., 2012b). The yellow has proved to be the most suitable for fibre production. Flax that is pulled too early -green – produces very fine but weak fibres. On the other hand, in overripe flax – brown – the stems are strong but brittle but produce too high a proportion of undesirable short fibres. When the flax is yellow, the fibres are long and supple, and therefore ideal for further processing. Therefore, the finest fibres are obtained by harvesting the crop following a full bloom with the stem and leaves green or at medium fibre fineness when half to a third of the seed bolls are yellow and brown with fully developed seeds (Butorac et al., 2009b; Muir and Westcott, 2003). In this period the flax stem turns yellow in the lower half, while the upper stem parts and capsules turn green yellow. In Croatia, flax harvest starts at the beginning of July.

2.1 Retting of flax steam

To obtain flax fibres for commercial use, flax steam undergo a process called retting. Retting, which is the separation of bast fibres from the core tissues, is very important and most difficult stage in flax fibre processing, as it affects quality and yield. There are many commonly known methods for plant fibre retting which allow for separation of the fibre from the woody part, and removal of non-cellulose components, such as pectin, hemicellulose, lignin, waxes and fasts (Sharma and Sumere, 1992; Sharma et al., 2011; Brunšek et al., 2014a).

After harvesting, the steams are usually kept either in the field or under water for 2 or 3 weeks, during which the pectinous substances that bind the fibre with other plant tissues are softened and degraded by micro-organisms. A quality of fibre is largely determined by the retting conditions and duration.

The two traditional types of retting are dew and water retting. With ***dew retting***, plant stems are cut or pulled up and left in the field to rot for over several weeks. Farmers monitor the process closely to ensure that the bast fibres separate from the inner core without much deterioration in quality. Moisture is needed for the microbial breakdown to occur, but then the weather must be dry enough for the steams to dry for bailing. The dew retting are depending on particular geographical regions that have the appropriate moisture and temperature ranges for retting, coarser and lower quality fibre than water retting and occupation of agricultural fields for several weeks and increase the risk of damage to the crop due to unsuitable weather during the retting period. Also during dew retting, the stems on the top of the swath take more moisture from rain or dew than those located on the bottom. They also dry faster between rains. This causes notable difference in retting between stems in top and bottom part of the swath. To improve the quality of obtained raw material the swath of straw is turned over – 2 or 3 times, depending primarily on the meteorological conditions during the retting period. If there is too much rain the straw may partially rot before retting is complete and this will weaken the fibres, thus reducing both quality and yield. On the other hand, if there is not enough moisture during the retting period, the straw will not ret sufficiently and the fibres will be damaged during scutching because of high content of shives and with residues of surrounding tissues still adhering to the fibre. Fibre ribbons will be characterized by a low divisibility and reduce fibre quality and yield (Franck, 2005; Bajpaj, 2012; Brunšek et al., 2014b).

In ***water-retting*** flax stems are submerged in rivers and lakes (recently in indoor pools and tanks), and anaerobic bacteria such as *Clostridium spp*., colonize the flax stems; produce pectinases and other enzymes to degrade pectin and other matrix compounds, thus freeing fibres from the core tissues (Franck, 2005). The retting

period varies from three days to a week, depending on the temperature of water. It must be carefully judged because under-retting makes separation difficult, and over-retting weakens the fibre. Water-retting results in high quality fibre which is long, strong and fine; therefore excellent for apparel and other textiles. On the other hand, the stench from anaerobic fermentation of the plants, extensive pollution of waterways, high drying costs, and putrid odour of resulting fibres resulted in a move away from anaerobic water-retting to dew-retting. However, the advantage of water- over dew retting is in that it is more controllable and avoids the risk of the crop being spoilt by inclement weather during the weeks that it lies on the ground (Franck, 2005; Sharma and Sumere, 1992; Bajpai, 2012; Brunšek et al., 2015; Akin, 2004).

An alternative that has had long-term consideration is ***enzyme-retting*** as a method to replace natural retting.

Enzyme-retting has been explored as an option that could be used in the world that would not only eliminate the reliance on favorable weather conditions, but also allow the industry to produce more consistent and high-quality fibres without the risks associated with uncontrolled natural retting.

Retting with enzyme mixture which consists of pectinases, hemicellulases and cellulases produced fibre fineness, strength, colour and waxiness comparable to the best water-retted fibre.

The advantages of enzyme-retting are offset by the additional costs and requirements of incorporating enzymes into a production environment that would include energy costs for drying, enzymes for retting, chemicals for maintaining processing conditions, and wastewater handling systems.

Nevertheless, it should be noted that retting alone is not sufficient to separate fibres from the straw and some mechanical processing is required, but the effort needed to perform the mechanical processing is significantly reduced on properly retted stalks. Either dew- or enzyme-retting must be integrated with mechanical processing for composite feedstock.

After retting and drying fibre bundles are still partly bound to wooden core from which they are separated by breaking and scutching. The next step is heckling whereby fibres are separated from short fibres, i.e. from tow. Shorter fibres and wooden core are baled. In several European countries short fibres are especially separated; they are used in making ropes, carpets and for production of yarn, generally in dry spinning system. Long fibres are used to make yarn (Butorac, 2009b).

3. Flax fibres

Flax as a renewable textile raw material is increasingly gaining in importance, the range of its use is spreading in different industry branches, resulting from the fact that all parts of the flax plant can be used. Besides for the conventional use, e.g. for casual and luxury clothing, interest for the use of flax fibres for technical textiles is increasing, whereby coarse flax fibres are gaining in importance, as a reinforcement in bio-composite materials. Although tow was less important in the past, nowadays it is considered to be a valuable ingredient of insulation materials (thermal/sound insulation), and it has found its use in the paper industry (banknotes). There is also interest in the use of wooden remains of the plant after the fibre extraction in the manufacture of interior car parts, storage containers, biodegradable containers, insulation of trusses and walls (Šurina et al., 2009; Sharma and Sumere, 1992).

Flax fibres are long and multinucleate cells without septum or partition. Fibres are specialised cells, having extreme length (average 2–5 cm) and secondary wall of large thickness (5–15 μm). They are gathered in bundles of one to three dozen cells that encircle the vascularcylinder (Fig.2).

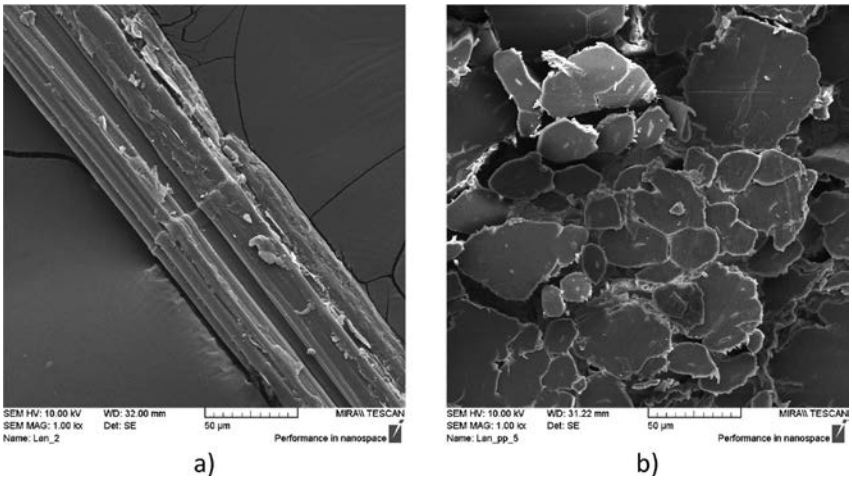


Fig. 2 – Flax fibres: a) technical bundle and b) cross section of elementary fibres

During plant growth, fibres develop in two main steps: cell elongation and thickening of secondary walls. Fibre-cell elongation (growing) is restricted to the top of the stem and is completed within few days. The fibre length is related to that of the internode, which contains the bundle and is maximum during the fast plant growth phase. The formation of secondary wall occurs centripetally, the first thickenings being observed immediately below the snap point. Most of the secondary wall deposits of cellulose happen after flowering, to completely fill the fibre at maturity of the capsule.The fibre secondary wall formation depends on mineral nutrition, espe-

cially on the calcium to sodium ratio and it is related to the water supply and light intensity. The diameter of the fibre cells considerably increases during the maturation of the capsule, in parallel with the wall thickening (Sharma and Sumere, 1992;Kelley et al., 2004; Morvan et al., 2003).

Plant fibres are constituted by three structural polymers (the polysaccharides cellulose, and hemicellulose and the aromatic polymer lignin) as well as by some minor non-structural components (i.e. proteins, extractives, minerals) (Sharma and Sumere, 1992). Cellulose is the basic building material of the flax fibre. In addition to cellulose, fibres contain a small proportion of hemicellulose and non-cellulose components: lignin, pectin, fats and waxes, water, pigments, minerals and ashes. The chemical composition of flax fibres depends on the cultivar, climatic conditions, cultivation of flax, the soil type, degree of ripeness and the steeping process. Data on the amount of individual constituents in flax fibres deviate in literature sources (Tab. 1) (Marques et al., 2010).

With a large amount of unidirectional cellulose microfibrils, flax fibre is a good model of a tubular structure of high tensile strength and low elasticity. Another specificity of flax fibre is that their cellulose microfibrils are embedded in non-cellulose polysaccharides with very low amount of lignin. The complex chemical structure of flax fibres certainly resembles the constitution of composite materials and may therefore be conditionally called a natural biocomposite. The cellulose molecules in flax fibres are bearing components; lignin is the strengtheners and pectin, hemicellulose, and the rest of the substances act as adhesives (Kelley et al., 2004; Morvan et al., 2003; Roach et al., 2011).

Table 1 – Chemical composition* (%) of flax fibres by various authors (Sharma and Sumere, 1992)

Cellulose	Hemicellulose	Pectin	Lignin	Fat and waxes	Authours
85	9	–	4	–	Couchman 1944
73	10	15	9	5	Overbeke 1949
71	19	2,0	2	1,7	Peters 1967
65	16	3,0	3	1,5	Haudek i Viti 1978
64	17	1,8	2	1,5	Anon. 1983
71,0	18,6	2,3	2,2	1,7	Bledzki 1996
64,1	16,7	1,8	2,0	1,5	Bledzki i Gassan 1999
66,14	16,24	2,14	2,0	1,09	Bhattacharya 2004

*In the table is not displayed numerical value of water, pigments, minerals and ash

Flax fibres are characteristic of a high level of non-uniform morphology which can be neither substantially affected nor removed by the production technology, but it noticeably affects fibre properties. Furthermore, the properties of flax fibres are also the result of the effect of different factors influencing the growth and development of flax plant like cultivars and agro-ecological cultivation conditions, applied agricultural techniques as well as fibre extraction from the stem of the flax plant and flax processing into fibres. During the acclimatization ability valorisation of cultivars, it is necessary to consider not only the plant properties but also the properties of the obtained fibres, such as length, fineness and tenacity, because they are closely related with the plant morphological characteristics such as plant height, technical stem length and stem thickness.

Closely related parameters, which have an impact on the quality of flax fibres, are agronomic, morphological and phonological plant properties such as root shape, length, thickness and stem colour, it's branching, leaf shape and retting procedure (Šurina et al., 2011). Elongated roots contain many primary roots or radicles and provide a uniform straight stem, which is high from 75 cm to 100 cm and thick from 1.5 to 2 mm, which is a prerequisite for high-quality flax. On the contrary, if plant height is smaller, fibre yield is lower and fibres are coarser. Early branching in the lower stem part will be the reason for lower fibre yield, and fibres are shorter and coarser. A quality flax stem has bright yellow leaves, it is lower, narrow and oblong. The stem, from which long and fibres with less lignin are extracted, can be recognized by yellowish colour with a greenish shade on the top, while the dark-coloured stem indicates exceeding harvest time, and fibres with a greater lignin content may be expected. It can also be asserted that each step from plant cultivation to fibre production has its own specific impact on fibre quality (Sharma and Sumere, 1992; Butorac, 2009b)

4. An interdisciplinary approach to flax research

With the aim of supporting a more intensive use of relatively quickly renewable raw materials and expansion of the use of natural fibres, interest in flax fibres has increased. A survey of journals covering the fields of textiles and agriculture, established an insufficient cooperation between researchers, especially focusing flax and its fibres. A one-sided approach in solving problems and reaching conclusions, with a lack of respect of facts concerning other professions, has been observed. Therefore, there is a need for an interdisciplinary approach to research, respecting agricultural and textile knowledge. In conformity with the world trend a complex project of revitalizing the production of staple flax fibre was initiated. Faculty of Textile Technology and Faculty of Agriculture researchers, University of Zagreb,

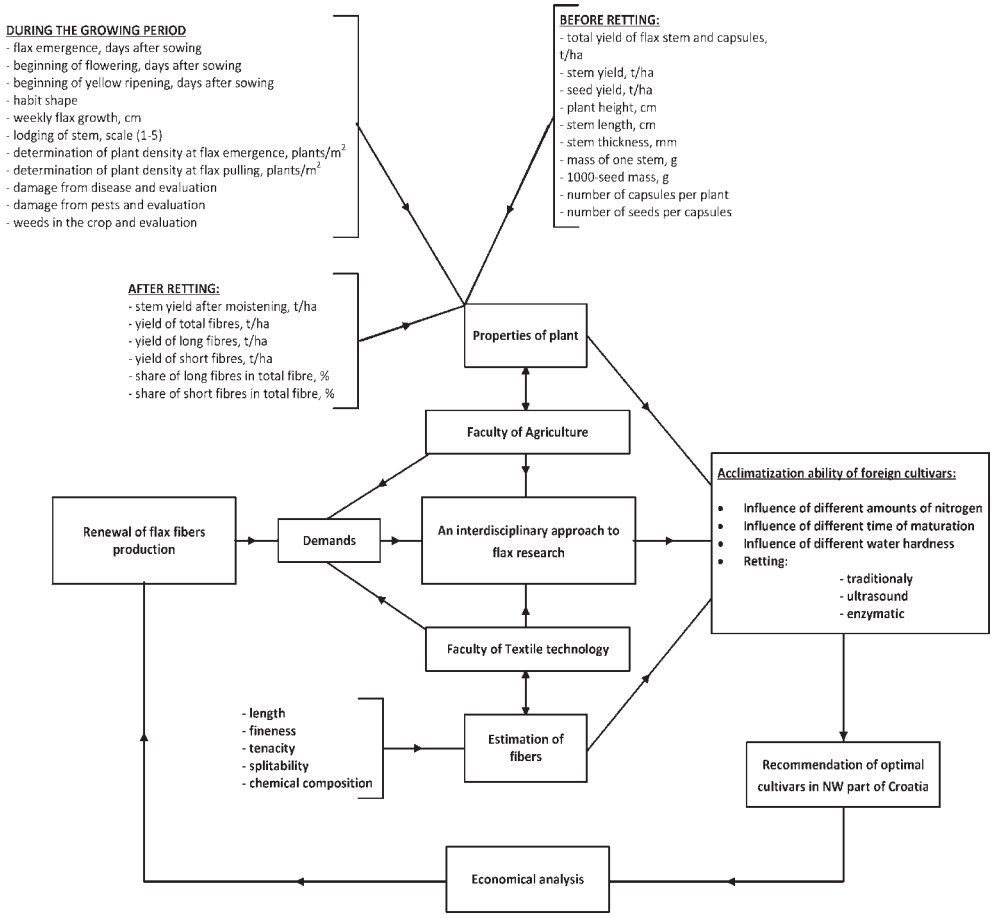


Fig. 3 – Project of the introduction of fibre flax in Croatia

have joined in a mission (Fig. 3), the aim of which is to renew the flax production in Croatia and define and improve the quality of flax fibres (Šurina et al., 2009; Andrassy et al., 2004; Butorac et al., 2004; Šurina et al., 2012c).

The fibre flax has traditionally been cultivated in most parts of Croatia but it has not own fibre flax cultivars. Therefore, it is necessary to introduce foreign flax cultivars, which have the greatest potential of the adaptation to Croatian climate conditions. Since flax grows in damp and moderately warm areas, in sensitive to high temperatures, hot and dry winds at the stage of intensive growth and flowering, it may be assumed that in newly formed climate conditions it comes to the reduction of its valuable agronomic properties and as a consequence to the reduction of textile technological fibre properties. Therefore, it is important to examine the acclimatization ability of the imported cultivars by determining agronomic properties of flax and textile technological fibre properties. Therefore, we intro-

duced cultivars from Western Europe (the greatest yield and quality) and these cultivars are less suitable for our climate (yields and quality are usually lower) (Šurina and Andrassy, 2013; Brunšek et al., 2014c).

4.1 Acclimatization ability of foreign fibre flax cultivars

As, is it shown at Figure 3, the main project objective is to identify the possibility of the introduction of foreign fibre flax cultivars to the continental lowland region of the north-western Croatia, i.e. to ensure raw materials for making ecologically valuable home-made products, widening the crop rotation of practised plant production systems and to ensure additional employment of family farms members. To achieve the set goal, it is necessary to get answers to fundamental questions:

– Influence of different amount of nitrogen
– Influence of different time of maturation
– Influence of different water hardness
– Ultrasound and enzyme retting

Methodology. Cultivar trials with fibre flax were carried out at the experimental field of the Faculty of Agriculture of Zagreb (45°49'26" N, 16°02'07" E), on anthropogenized eutric cambisol and of the College of Agriculture at Križevci (46°02'23" N, 16°54'62" E) on pseudoglay on level terrain (Fig.4). Both soils are silty loamy textures. Because of high powder content, they are prone to creating a crust. The soil in Zagreb has a weakly alkaline reaction, while the soil in Križevci is acid. At both locations it is richly supplied with phosphorus variable to the plant, moderately supplied with total nitrogen and it is weakly humic. In Zagreb it is well supplied with potassium accessible to the plant, while it is poorly supplied in Križevci. As part of the basic and pre-sowing seedbed preparation, 500 kg/ha of NPK fertilizer (7:20:30) was applied while 100 kg/ha of calcium ammonium nitrate (27%) was added with topdressing when the plants were 10 cm high. The trials were laid out according to the randomized complete block design with four replications. Sowing density was 2 500 germinable seeds/m^2. The main trial plot size was 10 m^2 (10 rows x 0.1 m row spacing x 10 m length).

The trials involved five cultivars: Viking (Cooperative Liniere de Fontaine Cany, France), Viola (Van de Bilt Zaden, Netherlands), Venica (Agritec, Czeck Rep.), Agatha, and Electra (Cebecco Seed, Netherlands). The fibre flax cultivars belong to the different vegetation ripening group. Viking belongs to the early vegetation ripening group, Venica to medium, Agatha to medium – late and Viola and Electra to half late (Butorac et al., 2011). All the cultivars have high resistance to the lodging and there are a high resistance to diseases (*Fusarium*). They have high yield of stem, high yield of fibre per hectare, high content of long fibre and good

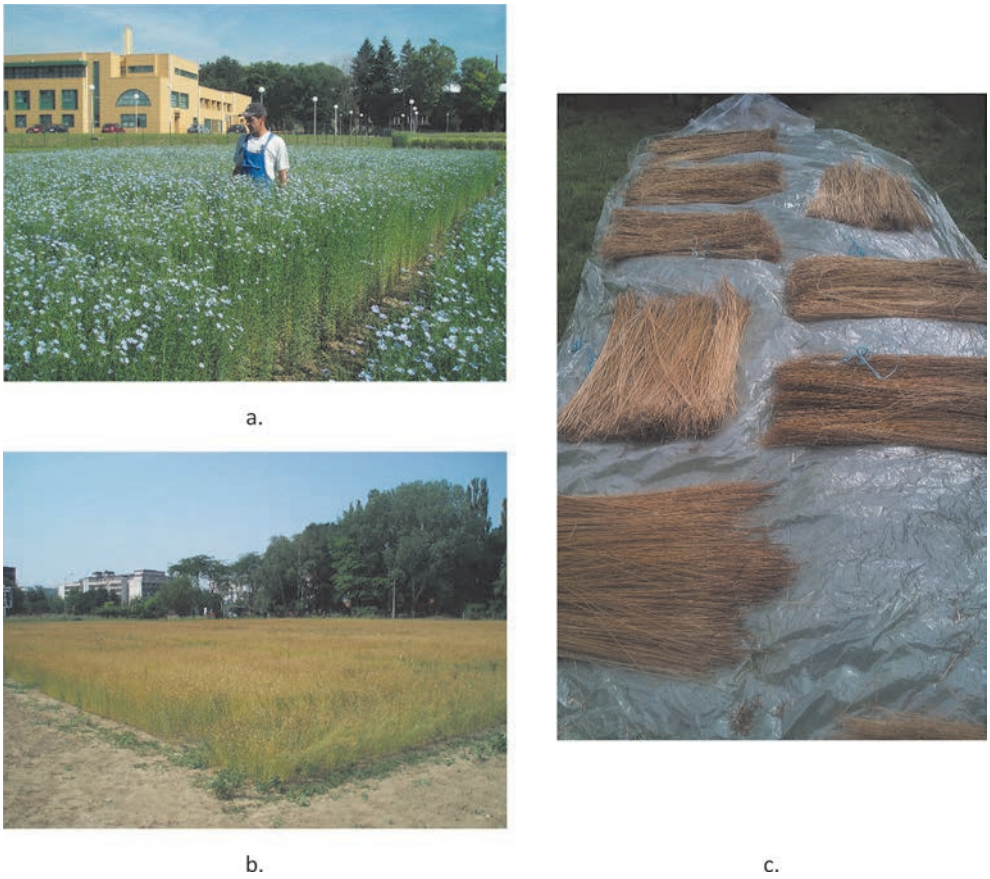


Fig. 4 – Flax: a) in the stage of flowering, b) in the stage of ripening and c) after retting

quality fibre. Only one cultivar of spring fibre flax, Viking (Šurina et al., 2011), is registered in Croatia. Fibre flax cultivar Viking was created in France in 1996. It distinguishes by early ripening, very high fibre yield per hectar and a share of total fibre up to 40%. It is very resistant to lodging and diseases (Butorac et al., 2006c).

Influence of different amount of nitrogen. Flax was fertilized with different nitrogen rates (0, 30, 60 and 90 kg/ha) at different development stages of plant. No nitrogen added in the first trial treatment. In the second fertilization treatment, all nitrogen added before sowing (30 kg nitrogen/ha). In the third fertilization treatment, 30 kg nitrogen/ha added before sowing, and 30 kg in a single fertilizer application at the average plant height of 10 cm; in the fourth fertilization treatment, 30 kg added before sowing, and 30 at the average plant height of 10 and 20 cm each.

Influence of different time of maturation. Influence of different time of maturation on flax and flax fibres properties was investigated by harvesting flax in three stages of maturity- green, yellow and brown ripe.

Influence of different water hardness. Retting of flax were carried out on a traditional way. Flax stems were subjected to biological maceration for 72 hours in hard hot water (32 °C) under controlled conditions. The investigation of influence of different water hardness during flax retting were carried out at the same way as traditionally retting but with three different water hardness – very soft, medium hard and hard. The water hardness is expressed in German degrees of hardness (°dH) and in ppm (determination by titration with 0.1 M HCl with methyl orange indicator) (Tab. 2).

Table 2 – Hardness of water used in retting

	Hard	Medium hard	Very soft
°dH	15.5 – 21	5.2 – 5.7	0.8 – 1.5
ppm	263 – 357	88 – 98	14 – 26

Ultrasound and enzyme retting. Because of problems with both water- and dew-retting, one of the objective is to develop physical retting which includes ultrasound retting of flax fibres and bio-innovative flax retting with enzyme. The use of ultrasound to extract fibres provides an environmentally friendly method. In this way, time of retting is shorter and there is no dependence on the weather conditions. This method allows to controlled water flow, temperature and time of the process and recycle used water. New enzymes have been recently evaluated to replace those currently used for retting. The enzymes remove the non-cellulosic compounds from the technical fibres, and at the same time are biodegradable and non-toxic, with no dependence on the weather conditions. The use of enzymes to extract fibres provides an environmentally friendly method toward developing reliable and sustainable agriculture using bio-based fibres of enhance quality.

Methods. During the vegetation period of flax beginning of flowering and early yellow maturity were monitored. Plant height was measured, prior to maceration, from the cotyledon node to the plant top, and the technical stem length from the cotyledon node to the first branch. Stem thickness was determined in the middle of the technical stem length using an electronic micrometer.

Also, the agronomic properties investigated were, steam yield, stem yield after retting, total fibre yield, share of total fibre, long fibre yield and share of long fibre. Stem yield was determined in the early yellow ripening stage.

After retting, stems were removed from the tank. They were dried at 60 °C for 30 hours and weighed. A scutching machine was used to separate straw (woody matter) from fibre, whereupon the yields of total and long fibres (using a set of hackling pins), and their respective share, were estimated (Fig. 5).

For the obtained fibres textile-technological properties were defined.

Fig. 5 – From flax to flax fibres after retting

Table 3 shows investigated fibre's parameter that are relevant for flax fibre's evaluation i.e. its quality.

Data of all properties studied were statistically processed by the analysis of variance (two-factor trial). Differences between mean values were analysed using Duncan's multiple range test (DMRT).

Table 3 – Performed measurements and methods

Parameters	Methods
Length of fibre	HRN ISO 6989:2003
Fineness Tenacity	HRN EN ISO 1973:2008, Vibroscop and Vibrodyn 400
Splitability	$A = \frac{F}{G}$ According to (Herzog, 1989) where: A–splitabillity, F–number of elementary fibres, G–number of fibres group (technical fibres)
Chemical composition: cellulose, hemicellulose, pectin, lignin, accompanying substances, fats and waxes	Common procedure for determining the chemical composition of plant fibres based on references (Herzog, 1989; Garner, 1967)
Used standards and regulations are adapted to testing technical flax fibres. Due to the non-homogeneity of flax fibres, the number of measurements was increased and determined according to statistical indications of the degree of reliability. Measurements of fibres properties were performed on conditioned samples.	

5. Conclusion

It can be expected that fibre flax production will be resumed in the lowland continental part of north-western Croatia by introduction of newly created, intensive (in terms of agricultural management and production) cultivars from Western Europe through valorization of their agronomic and textile-technological properties, and acclimatization ability.

From the agrotechnical, ecological and plant-production viewpoints, the inadmissibly narrow crop rotation would be extended; from the economic point of view, employment and incomes of a number of family farms would be increased (making homemade products of high quality), while the cultural aspect would involve dissemination of traditional values and culture.

Adequate selection of cultivars and production technology as well as processing would provide home-produced natural fibres for the textile industry, and raw materials for other branches of economy, with minimal environmental pollution.

One should not neglect the fact that the fibre production can play an important role in the existence of population and general improvement of the economic situation, among other things by export of those products, which nowadays possess much appreciated national characteristics integrated into valuable souvenirs.

References

[1] Akin, D. E. (2004). Linen most useful: Perspectives on structure, chemistry and enzymes for retting flax. ISRN Biotechnology (2013), pp. 1-23

[2] Andrassy, et al. (2004.). Resumption of the production of spinnable flax fibers. Tekstil Vol. 53 (8), 2004. str. 385-391

[3] Andrassy, M. et al. (2010). Influence of Agroecological Conditions on the Properties of Flax Fibres in Croatia, Book of Proceedings of 5th International Textile, Clothing and Design Conference – Magic World of Textiles. Dragčević, Zvonko (ed.), Faculty of Textile Technology, University of Zagreb. Dubrovnik, Croatia. 03.-08.10. 2010, pp. 32-37

[4] Bajpai, P., Biotechnology for Pulp and Paper Processing, (2012) Springer

[5] Brunšek, R. et al (2014a). Retting of flax and characterization of fibres. Book of Proceedings of the 14th World textile conference, Recep, E. (ed.), Uludag University, 26.-28.05. 2014., pp. 1-6

[6] Brunšek, R. et al: Influence of water hardness on retting and properties of flax fibers, Book of Proceedings of the 7th International Textile, Clothing & Design Conference Magic World of Textiles. Dragčević, Z., Hursa Šajatović, A., Vujasinović, E. (ed.). Zagreb, Faculty of Textile Technology, University of Zagreb, Dubrovnik, Croatia, 05.-08. 10. 2014. pp. 41-46, (2014b)

[7] Brunšek, R. et al: The revitalization of flax in Croatia. Tekstil Vol. 63 (1-2), 2014. pp. 49-58 (2014c)

[8] Brunšek, R. et al, Bio-innovative flax retting. Book of Proceedings of 8th Central European Conference on Fiber-Grade Polymers, Chemical Fibers and Special Textiles, Dekanić, T., Tarbuk, A., (ed.). Zagreb, University of Zagreb, Faculty of Textile Technology, Zagreb, Croatia, 16.-08.09. 2015. pp. 2-7

[9] Butorac, J. et al.: Possibilities of introduction foreign fiber flax cultivars in the lowland continental part of Croatia, Proceedings of the 3rd Global Workshop of the FAO European Cooperative Research Network on Flax and other Bast plants "Bast fibrous plants for healthy life", Kozlowski, R. (ed.), Banja Luka, October 2004, Institute of Natural Fibres, Poznan, 2004. pp. 1-9

[10] Butorac, J. et al. (2006a.). Estimation of agronomic and morphological traits of fiber flax varieties without and with topdressing by nitrogen. Sjemenarstvo Vol. 23 (5-6) 2006. str. 437-446

[11] Butorac, J. et al. (2006b.). The effect of plant density on some morphological and phenological traits of fiber flax varieties. Sjemenarstvo Vol. 23 (5-6) 2006. str. 447-456

[12] Butorac, J. et al (2006c). Some characteristic of European fiber flax varieties, Book of Proceedings of 41th Croatian and 1st International Symposium on Agriculture. Jovanovac, S. Opatija, February 2006. Poljoprivredni fakultet J.J. Strossmayera Osijek, Osijek, 2006. pp. 359-360

[13] Butorac, J. et al. (2009a). Estimation of agronomic and morphological traits of fiber flax varieties without and with topdressing by nitrogen. Sjemenarstvo Vol. 26 (3-4) 2009. str. 119-129

[14] Butorac, J., Fiber Crops, Kugler, Zagreb, Croatia (2009b)

[15] Butorac, J. et al. (2010a). The effect of nitrogen fertilization on fibre flax yield and share of fiber, Zbornik radova 45th Croatian and 5th International Symposium on Agriculture. Marić, S. (ed.), Opatija, Croatia, February 2010, pp. 681-685, Poljoprivredni fakultet J.J. Strossmayera Osijek, Osijek

[16] Butorac, J. et al. (2010b). Some Agronomic and Textile Properties of Flax Cultivated in Croatia (Krizevci), Book of Proceedings of 5th International Textile, Clothing and Design Conference – Magic World of Textiles. Dragčević, Zvonko (ed.), Faculty of Textile Technology, University of Zagreb. Dubrovnik, Croatia. 03.-08.10. 2010, pp. 42-47

[17] Butorac, J. et al (2011). The influence of vegetation ripening group of fiber flax cultivars on morphological and textile- technological traits, Book of Proceedings of 46th Croatian and 6th International Symposium on Agriculture, Pospišil, M. (ed.), Opatija, Croatia, February 2011., University of Zagreb, Faculty of Agriculture, 2011., pp. 723-727

[18] Butorac, J. et al (2012). The influence of stages of maturity on fiber flax yield and share of fiber. Book of Proceedings of 47th Croatian and 7th International Symposium on Agriculture, Pospišil, M. (ed.), 13.-17.02. 2012., Opatija, Croatia, pp. 464-468

[19] Butorac, J. et al. (2014). The influence of nitrogen on the agronomic traits of fibre flax cultivars. Romanian Agricultural Research Vol. 1 (31), 2004. pp. 1-7

[20] Franck, R. R. Bast and other plant fibres. Cambridge, Woodhead Publishing Limited. (2005)

[21] Garner, W., Fibres-In *Textile laboratory manuel*, Heywood Books, London, (1967), pp. 52-113

[22] Harris, J., 5000 Years of Textiles. 2nd edition. London, The British Museum Press. (2004)

[23] Herzog, W. (1989). Objektive Qualitats prufung von Flachs, Melliand Textilberichte Vol. 63 (1) 1989. pp. 7-11

[24] Jeny, N. H. A not very famous museum-but not without interest. Izvor: http://www.carment-ablog.com/2015/02/06/not-famous-museum-not-without-interest/ (Datum zadnjeg pristupa: 03. listopada 2015.)

[25] Kelley, S. S. et al (2004). Rapid analysis of the chemical composition of agricultural fibers using near infrared spectroscopy and pyrolysis molecular beam mass spectrometry. Biomass and Bioenergy Vol. 27 (1) 2004. pp. 77 – 88

[26] Marques, G. et al (2010). Evaluation of the Chemical Composition of Different Non-Woody Plant Fibers Used for Pulp and Paper Manufacturing. The Open Agriculture Journal Vol. 4, 2010. pp. 93-101

[27] Morvan, C. et al (2003). Building flax fibres: more than one brick in the walls. Plant Physiology and Biochemistry Vol. 41 (11-12) 2003. pp. 935–944

[28] Muir, D. & Westcott, N. D., Flax: The Genus Linum (Medicinal and Aromatic Plants – Industrial Profiles), CRC Press, Canada, (2003)

[29] Paskovic, F. (1954.). Flax in Croatia and measures for its improvement. Tekstil Vol. 4 (4), 1954. str. 267-273

[30] Paskovic, F. (1965.). European problems in flax production. Tekstil Vol. 15 (7), 1965. str. 254-515

[31] Preasted, J. H., Ancient records of Egypt. Volume 1. Chicago, The University of Chicago Press. (1906)

[32] Roach, M. J. et al (2011). Development of Cellulosic Secondary Walls in Flax Fibers Requires b-Galactosidase. Plant Physiology Vol. 156, 2011. pp. 1351–1363

[33] Sharma, H.S.S & Van Sumere C.F., The Biology and Processing of Flax- M Publications. Belfast, Northern Ireland. (1992)

[34] Sharma, H.S.S., et al (2011). An economic and eco-friendly approach for degumming of flax fibres. World Journal of Microbiology and Biotechnology Vol. 27 (11) 2011, pp. 2697-2701

[35] Šurina, R. et al. (2009). Flax – plant and fibres through century. Tekstil Vol. 58 (12) 2009. pp. 625-639

[36] Šurina, R. et al (2011). Dependence of fiber properties on the properties of the flax fibers. Tekstil Vol 60 (2-3) 2011. pp. 87-101

[37] Šurina, R. et al (2011). Viking – standard cultivar of fiber flax in Croatia. Book of Proceedings of 4th International scientific-professional symposium Textile Science and Economy, Ujević, D., Penava, Ž. (ed.), University of Zagreb, Faculty of Textile Technology, Zagreb, Croatia, 26. January 2011. pp. 87-90

[38] Šurina, R. et al (2012a). An interdisciplinary approach to flax research. 12th AUTEX World Textile Conference – Innovative textile for high future demands, Mijović, B., Ujević, D., Petrak, S., Grancarić, A.M., Glogar, M. I., Salopek Čubrić, I. (ed.). Zagreb, Tiskara Zrinski d.d., Čakovec, 13.-15. 05. 2012., pp. 181-186

[39] Šurina, R. et al (2012b). Influence of maturation on the chemical composition of flax fibres, Book of Proceedings of the 6th International Textile, Clothing & Design Conference Magic World of Textiles. Dragčević, Z., Hursa Šajatović, A., Vujasinović, E. (ed.). Zagreb, Faculty of Textile Technology, University of Zagreb, Dubrovnik, Croatia, 07.-10.10. 2012., pp. 110-115

[40] Šurina, R. et al. (2012c). Three-years of flax and flax fibres research, Book of Proceedings of the 6th International Textile, Clothing & Design Conference Magic World of Textiles, Dragčević, Z.; Hursa Šajatović, A.; Vujasinović, E. (ur.), Zagreb, Faculty of Textile Technology, University of Zagreb, 07.-10. 2012. pp. 116-121

[41] Šurina, R. & Andrassy, M.: Renewal of flax fibres production in Croatia. 6th Scientific-professional symposium Textile Science and Economy,, Penava, Ž., Gudlin Schwarz, I. (ed.). Zagreb, Croatia 2013, pp. 11-16

Scientific paper

Predicting Process Capability Index in Early Stage of Manufacturing System Design

Tonči Mikac*, Sandro Dobovičekk*

Faculty of Engineering, University of Rijeka

Product quality is one of the most influential decision-making factors in process of manufacturing system design. Expected process capability is one of the variables for predicting configuration capability index (CCI) for newly designed manufacturing systems and it is based on historical dataset from similar processes in the past. Logical procedure for pretreatment of historical data before calculating of expected process capability indices are proposed in this article. Application of the procedure is shown in the example from automotive industry.

Key words:

Manufacturing system design, product quality, statistical process control, probability distributions

1. Introduction

Quality is one of the most important decision factors when choosing between the offered products or services. Thereby, it is irrelevant whether the customer is an individual, small or a large corporation or financial institution. The concept of quality and improvements of the key factors that influence the quality can lead to business success, growth and increase competitiveness.

Design of manufacturing system is a very complex process in which designers seeks to include all influential factors. The problems that arise in the manufacturing system design are therefore subject to constant scientific and practical research.

*tmikac@riteh.hr, sdobovicek@riteh.hr

The main goal of that process is to design the manufacturing system which will ensure all the required system specifications such as required product quality, sufficient process efficiency, easy process handling etc.

Decisions made during the manufacturing system design can affect the overall capability of processes and therefore affect quality of the final product. To estimate a value for the overall capability of processes at the early system design (concept) stage, the availability of process capability databases is essential (Nada et al., 2006.). Process capability databases which are developed by manufacturers for their processes and machines, use historical data to predict the future process capability index for a products based on the specific manufacturing process used to produce them. The process capability indices obtained using historical databases is based not only on the particular manufacturing process but also on other factors that influence the process such as the material type, size of the feature, and tolerances.

Nada et al. (2006.) are assumed that all processes follows a normal distribution. This assumption facilitates the conversion of process sigma level into the associated process yield value using tables for the area under standard normal distribution curve. Various types of influences that operate within a relatively simple production systems lead to the fact that, in these processes, a complex system of governing actions on the output parameters exists. This sometimes leads to the fact that processes don't follow a normal distribution.

In real life environment, characteristics which are expected to be normally distributed often are not. The causes of this should not seek in appear of non-normal influences. Examples are many, and the causes may be from a completely different nature. In mass production, various operators using multiple identical machines aim to make identical products, which all must meet set thresholds in characteristics such as shaft diameter and sheet thickness. Accordingly, disparities in operators' skills and machine performance are worth consideration when calculating process capability. The overall histogram can signal bimodal distribution – the mixture of two normal distributions. The root cause for such a phenomenon is the use of two spray guns (Harris et al., 2009.). Bimodal distribution is commonplace in processes where two pieces of manufacturing equipment are employed. Care should be exercised when multiple identical machines are allocated to produce the same parts because distribution may appear non-normal. As a result, it is difficult to troubleshoot if data is not categorized by machines and operators.

As Levinson (2011.) noted, a normal distribution is far more common in statistical textbooks than it is in real-world processes, and untold grief results from the unquestioning assumption that all manufacturing processes follow the bell curve. Therefore sigma levels (process capability indices) predictions for newly designed

manufacturing systems which are based on historical data should be performed in systematic and logical approach described in this article.

2. Non-normal distribution scenarios

Selection of the mathematical distribution that best fit the observed group of data can be very important for making correct prediction of process sigma level. Typical distributions are Uniform, Normal, Exponential, Weibull, Gamma, Loglogistic, Lognormal and others. The Uniform distribution has a rapidly increasing hazard function. The Normal distribution has a gradually increasing hazard function. The exponential distribution has a constant hazard distribution. The Weibull and Gamma distributions have decreasing hazard functions for shape parameters between 0 and 1, constant hazard functions for shape parameter 1 and increasing hazard function for parameters larger than 1. Exponential distribution is a special case of the Weibull and Gamma distributions. The Loglogistic distribution has a decreasing hazard function when the shape parameter is less than 1 and has one mode when the shape parameter is larger than 1 (Cox and Oakes, 1984.).

When the natural distribution of a dataset is non-normal, there is several ways to determine process capability index and process sigma level. First, data can be transformed so that they follow the Normal distribution. A second approach is to find a non-normal distribution that fits the data. Many non-normal distributions can be used to model a response, but if an alternative to the normal distribution is going to be viable, the Exponential, Lognormal, and Weibull distributions usually work. (Johnson, 2007.).

2.1 Data transformations

Data transformations are commonly-used tools that can serve many functions in quantitative analysis of data, including improving normality of a distribution and equalizing variance to meet assumptions and improve effect sizes, thus constituting important aspects of data cleaning and preparing for statistical analyses. There are as many potential types of data transformations as there are mathematical functions. Some of the more commonly-discussed traditional transformations include: adding constants, square root, converting to logarithmic (e.g., base 10, natural log) scales, inverting and reflecting, and applying trigonometric transformations such as sine wave transformations. However, a cautionary note is in order. While transformations are important tools, they should be utilized thoughtfully as they fundamen-

tally alter the nature of the variable, making the interpretation of the results somewhat more complex (Osborne, 2010.).

Generally, if data transformations are applied, it is important to understand the context of data. Data transformations are considered reasonable only if the meaning and purpose of data analysis is taken into account. Software transformation of data cannot be done automatically without the recognition of context.

Osborne (2010.) presents Box-Cox transformation (BCT) procedures to researchers as a potential best practice in data cleaning. The Box-Cox transformation is a particularly useful family of transformations. It is defined as (NIST/SEMATECH e-Handbook of Statistical Methods, 2013.):

$$T\left(Y\right)=\left(Y^{\lambda}-1\right)/\lambda.$$

The Box–Cox transformation is usually called "the power transformation" because the data are transformed by raising the original measurements to a power of lambda, so it is limited to a non-negative data values. Second commonly used type of data transformation is Johnson's transformation.

2.2 Finding suitable non-normal distribution

The Weibull distribution is the most widely distribution used in industry. Two main reason for popularity of Weibull distribution are that it has a simple expressions and closed forms to model the probability density function (PDF), reliability function, cumulative distribution function (CDF) and hazard function (Liu, C.-C., 1997.). The Weibull probability density function is (NIST/SEMATECH e-Handbook of Statistical Methods, 2013.)

$$f\left(x\right)=\frac{\gamma}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{(\gamma-1)}\exp\left(-\left(\left(x-\mu\right)/\alpha\right)^{\gamma}\right)\;\; x\geq 0;\gamma>0$$

where γ is the shape parameter, μ is the location parameter and α is the scale parameter, figure 1(a). Special case where $\mu=0$ and $\alpha=1$ is called standard Weibull distribution. Frequently, the location parameter μ is not used, and the value for this parameter can be set to zero. When this is the case, the PDF equation reduces to that of the two-parameter Weibull distribution.

In engineering, Lognormal distribution plays an important role in statistically predicting the fatigue life of mechanical products. The lifetime of products under any stress levels follows lognormal distribution (Li et al., 2015.). Since this includes

most, if not all, mechanical systems, the lognormal distribution can have widespread application. Consequently, the Lognormal distribution is a good companion to the Weibull distribution when attempting to model these types of units. The general formula for the PDF of the Lognormal distribution is (NIST/SEMATECH e-Handbook of Statistical Methods, 2013.)

$$f(x) = \frac{e^{-\left(\left(\ln\left((x-\theta)/m\right)\right)^2/\left(2\sigma^2\right)\right)}}{(x-\theta)\sigma\sqrt{2\pi}} \quad x \geq 0; m, \sigma > 0$$

where σ is the shape parameter (and is the standard deviation of the log of the distribution), θ is the location parameter and m is the scale parameter (and is also the median of the distribution), figure 1(b). If $x = \theta$, then $f(x) = 0$. The case where $\theta = 0$ and $m = 1$ is called the standard Lognormal distribution. The case where θ equals zero is called the 2-parameter lognormal distribution.

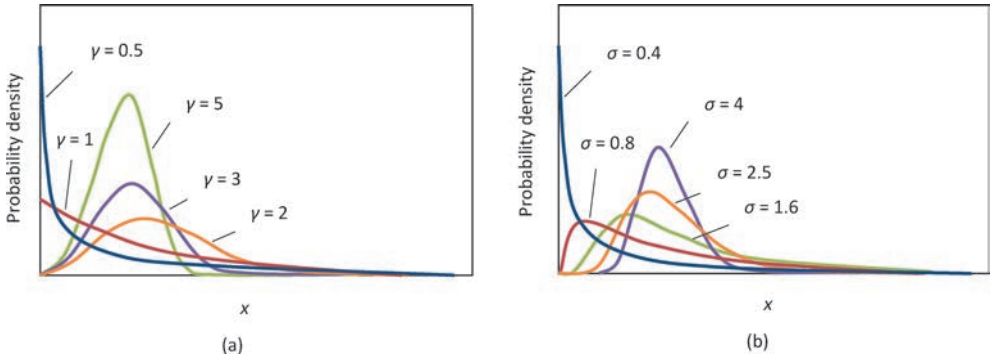


Fig. 1 – Probability density distributions: Weibull (a) and Lognormal (b)

Although the Weibull and Lognormal distributions are the most commonly used non-normal distributions, there are a number of known mathematical distributions in which set of observed data can be approximated by. Common way to find a suitable distribution is use of statistical hypothesis presented in a way that can be valued with statistical-analytical procedures.

A statistical hypothesis is a mathematical expression that represents the basis for the statistical test calculation. Hypothesis test is a statistical procedure that determines whether and how reliable the available data support the hypothesis set. Hypothesis testing and significance testing is basically the process of quantifying the impressions on these hypotheses. The null hypothesis, H_0, represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. The significance level α of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H_0, if it is in fact true.

The Anderson-Darling test is widely used to test if a sample of data came from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than does the K-S test. The K-S test is distribution free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution.

The sequence of actions for hypothesis test is: (1) set up the null hypothesis and alternative hypothesis, (2) significance level α choice, (3) sample data collection, (4) calculating the value of a specific statistical test results for the null hypothesis H_0, (5) compare the results of the statistical test with the values of specific probability distributions for a given test, (6) result interpretation in terms of probability (P-value). Significance level α should be selected with regard to possible consequences, usually 0.05 or 0.01. Hypothesis testing can verify whether the observed distribution fits a defined, non-normal or normal distribution. If P-value exceeds the significance level α, hypothesis is accepted, otherwise rejected.

2.3 Applying logical procedure

A large number of mathematical and statistical models for calculation of process capability that are available can be confusing and not always give useful results. Proposed logical procedure summarizes available possibilities when observed historical process data set received from historical database shows signs of non-normality, figure 2. It should be noted that the awareness of the expected form the distribution of probability and context data are extremely important. It is desirable that the observed data have traceability marks with the time and machine of part creation, as well as time of making measurements.

Logical approach includes pre-processing of data which allows more accurate conclusions about the future state of the process. Logical procedure has two possible ends: predicting process capability as for normal distribution and predicting process capability as for suitable non-normal distribution. All required tools are available in statistical software like widely applied Minitab. Minitab is often used in conjunction with the implementation of Six Sigma, CMMI (Capability Maturity Model Integration) and other statistics-based process improvement methods. It includes various types of data transformations, numerous mathematical distributions with a possibility of testing hypothesis. This possibility makes Minitab a suitable tool for application of logical procedure.

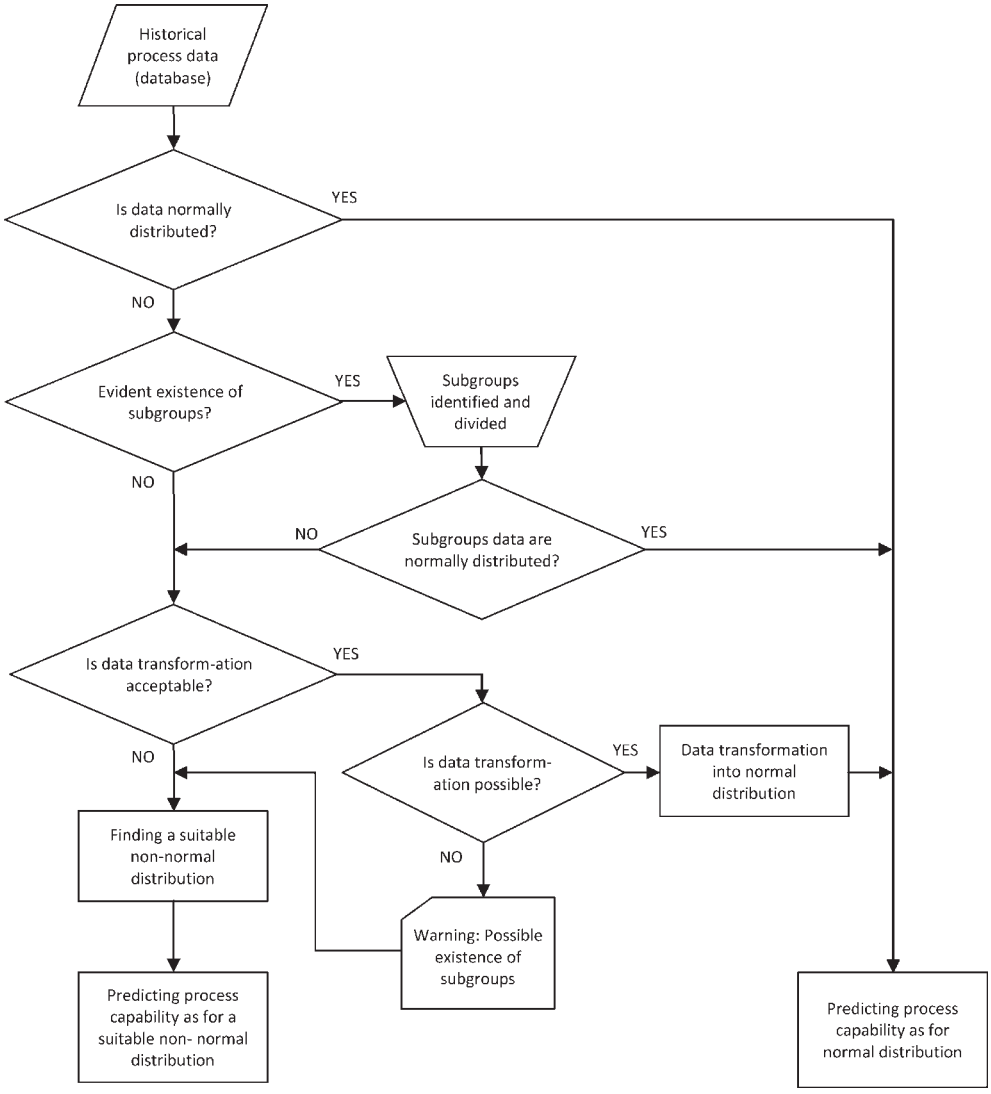


Fig. 2 – Logical procedure for predicting process capability from historical process data

The input (historical process data set) obtained from a database is first subjected to a normality test with a certain significance level. In case that normality test is positive, the process capability index and belonging process yield are calculated using tools for the normal distribution. If input set of data fails the normality test, and especially if normal distribution is anticipated, it is wise to check if there is more sub-groups in data. If subgroup existence is evident, subgroups should be separated. Re-verification of the hypothesis of normal distribution at each of them separately should be made.

If observed data is expected to be normally distributed but it is not, data transformation may be possible solution. Impossibility of data normalization can be a warning of subgroups existence inside the observed data and signal for data revision.

Finding suitable non-normal distribution involves interpretation of P-value for a goodness of fit test, such as Anderson-Darling, when using Individual distribution identification. A P-value less than α suggests that the data do not follow that distribution.

As an example, let's assume that new manufacturing system is designed. Product is a new model of gear lever mechanism. In the early phase of system design, various alternative solutions of processing technology are considered. Those various solutions can be graded according to the predicted process capabilities. To predict process capabilities for one solution, a historical process data for similar products manufactured similarly in the past (figure 3) should be used.

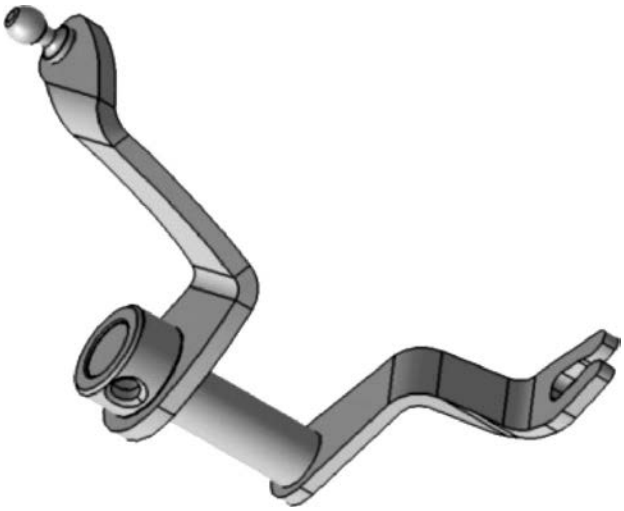


Fig. 3 – Part of a gear lever mechanism

Common case is that the work piece, whose production requires more types of technology, takes on more production sites – locations. The production process of similar part in the past included several different technologies and operations, and complete production process takes place at several locations. Involved technologies were metal forming, machining, surface protection and mounting. The quality of the final product depended on the overall production process where formation of subgroups inside the final measured data are likely.

Table 1 – Important dimensions

Index	Dimension	Expected distribution
D1	length 29±0.3	normal
D2	perpendicularity 0.6 max.	non-normal
D3	parallelism 0.6 max.	non-normal
D4	length 47.48±2	normal
D5	length 83.3±1.5	normal
D6	length 90±0.5	normal
D7	length 90±0.5	normal
D8	length 69.2±0.35	normal

The product has a total of eight functionally defined important dimensions (special characteristics) just like previous model manufactured in the past, table 3. For two dimensions with a boundary limit (perpendicularity D2 and parallelism D3), distribution is expected to be skewed or non-normal. For the simplicity, only two important dimensions are considered: one that is expected to be normally distributed (D8) and one that is expected to be non-normally distributed (D3).

3. Results and Discussion

Anderson-Darling (A-D) normality test showed P-values for both distributions <0.05 (chosen significance level is $\alpha = 0.05$) which means that none of the observed distribution cannot be accepted as normal.

It can be concluded that the dimension D3 is distributed as expected, figure 6. Histogram of D8 shows signs of sub-grouping, figure 7. Although there is no precise mathematical solution to the identification and separation of these data sets, this indicates an existence of subgroups.

Box-Cox transformation of D3 finds the optimal values of λ in the $\lambda = -0.02$ while same method transformation does not find the optimal λ in the range $-5 < \lambda < 5$ for D8. That was expected due the lack of such methods in cases when the distribution contains several sub-groups (camel hump shape). Johnson transformation indicates as a more powerful tool in both cases. Normality test of D8 after Johnson transformation shows $\mathrm{P-value} = 0.936$ which indicates that the null hypothesis of normal distribution is accepted.

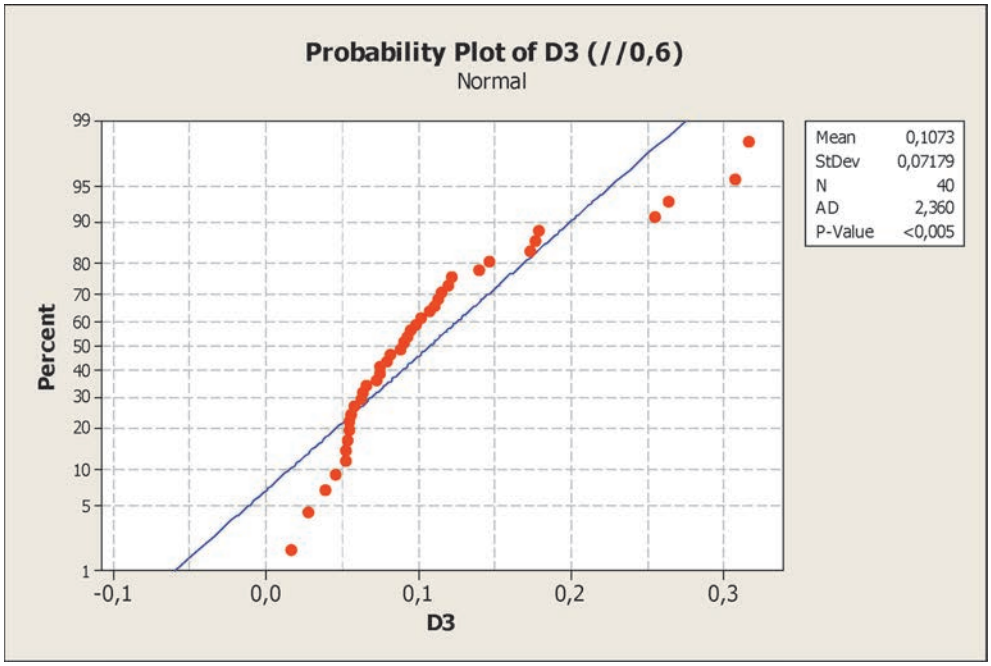


Fig. 4 – Probability plot of D3

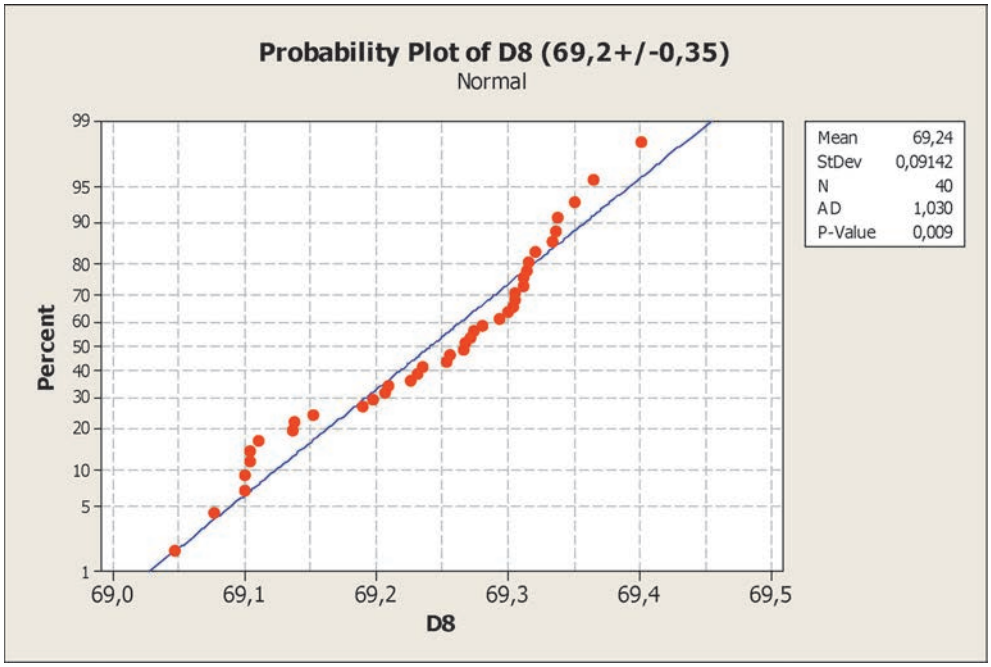


Fig. 5 – Probability plot of D8

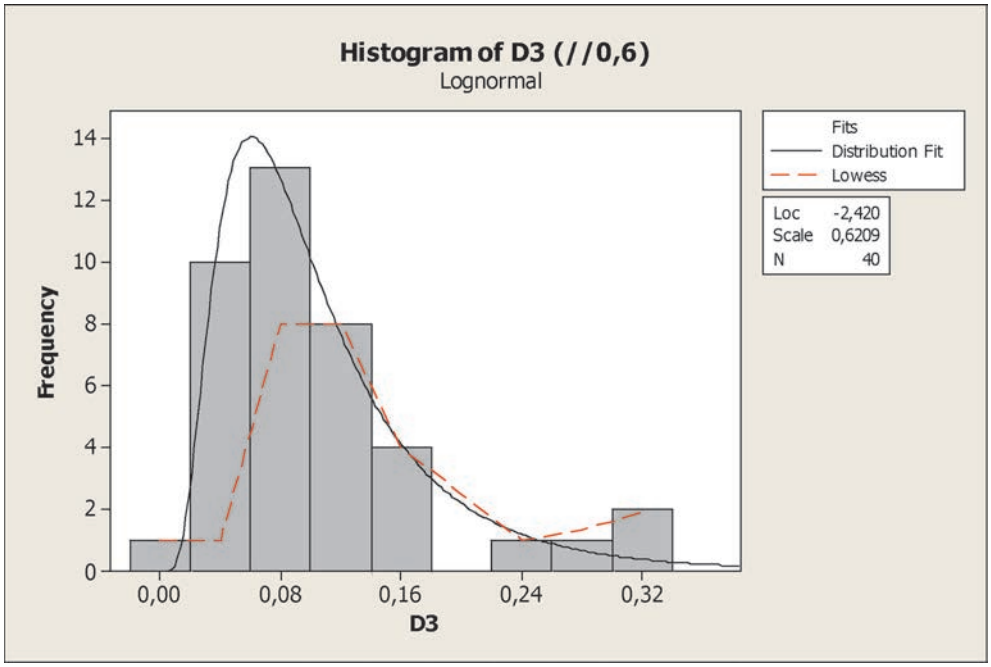


Fig. 6 – Histogram of D3

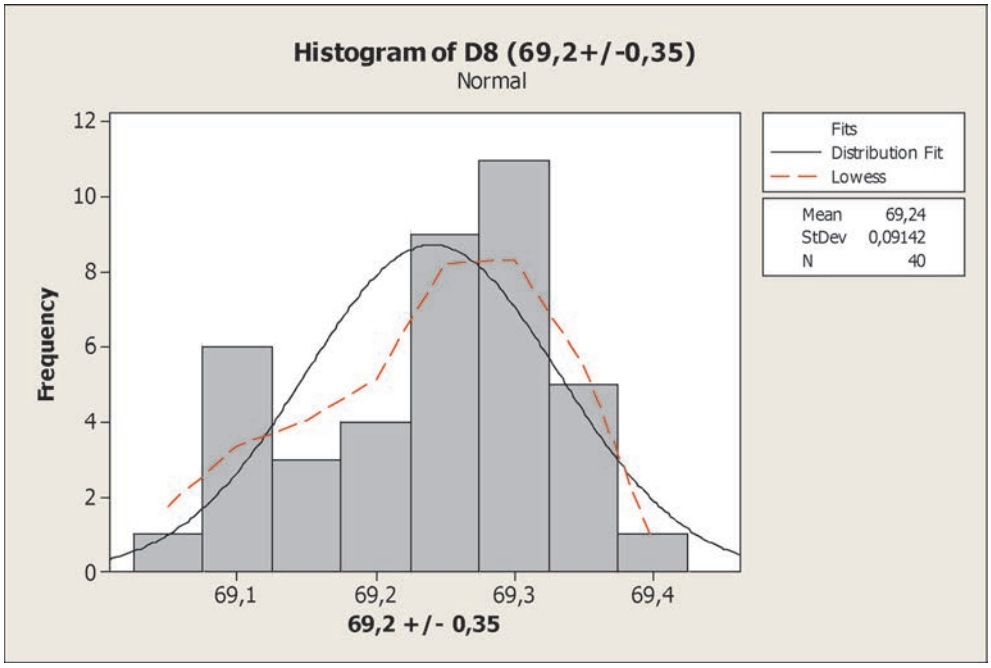


Fig. 7 – Histogram of D8

Considering the context of data that defines the distribution of characteristic D3, it is useful to calculate process capability on the non-transformed data. Prerequisite for such calculation is finding a non-normal distribution that fits the observed distribution. For D3 the test shows that for the significance level $\alpha = 0.05$ following distributions can be accepted: Lognormal ($P = 0.517$), 3-parameter Weibull ($P = 0.070$), Gamma ($P = 0.082$) and Loglogistic ($P > 0.250$), table 2.

Table 2 – Goodness of fit test and distribution parameters for characteristic D3 (Minitab)

Distribution	A-D	P-value	LRT P	Location	Shape	Scale	Threshold
Normal	2.360	<0.005		0.1073		0.07179	
Box-Cox transformation	0.322	0.517		–2.41985		0.62092	
Lognormal	0.322	0.517		–2.41985		0.62092	
3-Parameter Lognormal	0.328		0.815	–2.37907		0.58821	–0.00309
Exponential	3.542	<0.003				0.10730	
2-Parameter Exponential	2.256	<0.010	0.001			0.09262	0.01468
Weibull	1.052	<0.010			1.65390	0.12101	
3-Parameter Weibull	0.709	0.070	0.049		1.36800	0.10059	0.01551
Smallest Extreme Value	4.026	<0.010		0.14726		0.09004	
Largest Extreme Value	0.746	0.047		0.07793		0.04544	
Gamma	0.685	0.082			2.81915	0.03806	
3-Parameter Gamma	0.543		0.280		2.03494	0.04668	0.01232
Logistic	1.503	<0.005		0.09549		0.03570	
Loglogistic	0.237	>0.250		–2.43112		0.34188	
3-Parameter Loglogistic	0.222		0.773	–2.48587		0.36264	0.00435
Johnson Transformation	0.165	0.936		0.09058		0.95918	

The Likelihood Ratio Test (LRT P) rates significance of 3-parameter distribution from a 2-parameter distribution. If the P-value is less than α, then the improvement by using a 3-parameter distribution (instead of a 2-parameter distribution) is large enough to be statistically significant. Therefore, the 3-parameter lognormal distribution ($LRTP = 0.815$) and 3-parameter gamma distribution ($LRTP = 0.082$) can also be accepted.

Process capability analysis of characteristic D3 for each distribution with a P-value greater than a given significance level ($\alpha = 0.05$) results with a different capability indices C_{pk}, table 3. This shows the importance of this step where decision on the

selected distribution can be influential on prediction of future process capability for similar process. Johnson transformation of distribution D8 increases predicted capability indices P_{pk} which can be significant for predicting process capability, table 4. Transformation can be justified with expected normal distribution for this dimension.

Table 3 – Predicted process capability for characteristic D3 (Minitab)

Distribution	LB	USL	Target	Cpk
3-Parameter Lognormal	0	0.60	0	1.14
3-Parameter Weibull	0	0.60	0	1.57
3-Parameter Gamma	0	0.60	0	1.50
3-Parameter Loglogistic	0	0.60	0	0.62
Box-Cox transformation	–6.91*	–0.51	–6.91	1.04
Johnson transformation	–2.49	2.59	–2.49	0.87

*LSL

Table 4 – Predicted process capability for characteristic D8 (Minitab)

Distribution	LSL	USL	Target	Cpk	Ppk
Normal	68.85	69.55	69.200	1.84	1.13
Johnson transformation	–	–	–0.543	–	1.57

4. Conclusions

In the manufacturing system design process, prediction of process capability can be influential factor when choosing between several systems scenarios. Predicting process capability is possible if historical data for similar manufacturing process exists. But even if this kind of data exists, finding adequate calculation method for determination of process capability indices can be challenging.

Original contribution of this paper is expressed through setting up a network of logical questions and mathematical hypothesis, which together form the logical process to determine the process capability. In our research we came to the conclusion that the manufacturers who maintain their own process capability databases, often ignore data context. Online SPC applications that gather the data in real time and calculate process capability were usually programmed as a simplified, two-step process calculators. In first step, historical data are subjected to the normality test. Second step then is, depending on the result of normality test, to calculate process

capability index: for normal distribution in case when normality test is positive, or for Weibull distribution when normality test failed. Such simplified procedure leaves no opportunities for designers to further understand the nature of historical data as proposed logical procedure does.

Proposed logical procedure summarizes possibilities in case of data non-normality and helps the designers to interpret historical process data correctly. Benefit of this approach is reflected in the detection of possible cases where observed historical data set which defines a state of the process is non-normally distributed, whether the non-normality is expected or not. In cases where the expected normal distribution of the historical data set is not evident because of the existence of subgroups, this approach is lacking solution for afterward subgroup separation. This should be solved by proper workpieces marking and traceability in the manufacturing process (eg. date, shift, time, machine, tool, operator etc.). If adequate traceability is provided, proposed logical procedure can help the designers to interpret historical data and correctly predict process capability. Concrete implementation of this logical procedure presupposes the availability of some of the available software solutions like industry widely accepted Minitab or similar.

References

[1] Harris. M. A., Mynors D. J., Wang C. J.: A Low volume production process capability analysis, Applied Mechanics and Materials 16(19), 1038 – 1042, 2009.

[2] Johnson, L.: Modeling Non-normal data using statistical software, R&DMagazine, 26-27, www.rdmag.com, 2007.

[3] Levinson, W. A.: When the bell curve doesn't fit, Part 1 (2011.) Source: http://www.qualitydigest.com/inside/quality-insider-article/when-bell-curve-doesnt-fit-part-1.html (last accessed: October 14, 2015)

[4] Li, X., Pengfei, G., Fuqiang, S.: Acceptance sampling plan of accelerated life testing for lognormal distribution under time-censoring, Chinese Journal of Aeronautics 28(3), 814-821, 2015.

[5] Liu, C.-C.: A comparison between the Weibull and Lognormal models used to analyse reliability data, Department of Manufacturing Engineering and Operations Management, University of Nottingham, 1997.

[6] Nada, O. A., ElMaraghy, H. A., ElMaraghy, W. H.: Quality prediction in manufacturing system design, Journal of Manufacturing Systems 25(3), 153-171, 2006.

[7] NIST/SEMATECH e-Handbook of Statistical Methods (2012) Source: http://www.itl.nist.gov/div898/handbook/ (last accessed: October 14, 2015)

[8] Osborne, J. W.: Improving your data transformations: Applying the Box-Cox transformation, Practical Assessment, Research & Evaluation 15(12), ISSN 1531-7714, 2010.

Selection of Previously Published Papers by the Members of the Croatian Academy of Engineering

doi: 10.1016/j.jsbmb.2011.03.015

Demasculinization and Feminization of Male Gonads by Atrazine: Consistent Effects Across Vertebrate Classes*

Tyrone B. Hayes[a,*], Lloyd L. Anderson[b], Val R. Beasley[c], Shane R. de Solla[d], Taisen Iguchi[e], Holly Ingraham[f], Patrick Kestemont[g], Jasna Kniewald[h], Zlatko Kniewald[h], Valerie S. Langlois[i], Enrique H. Luque[j], Krista A. McCoy[k], Mónica Munõz-de-Toro[j], Tomohiro Oka[l], Cleida A. Oliveira[m], Frances Orton[n], Sylvia Ruby[o], Miyuki Suzawa[f], Luz E. Tavera-Mendoza[p], Vance L. Trudeau[q], Anna Bolivar Victor-Costa[m], Emily Willingham[r]

[a]Laboratory for Integrative Studies in Amphibian Biology, Molecular Toxicology, Group in Endocrinology, Energy and Resources Group, Museum of Vertebrate Zoology, and Department of Integrative Biology, University of California, Berkeley, CA 94720, USA

[b]College of Agriculture and Life, Sciences Professor of Biomedical Sciences, College of Veterinary Medicine Iowa State, University of Science and Technology Department of Animal Science, 2356 Kildee Hall, Ames, IA 50011-3150 USA

[c]Envirovet Program in Wildlife & Ecosystem Health, Department of Comparative Biosciences, College of Veterinary Medicine, University of Illinois at Urbana-Champaign, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA

[d]Wildlife and Landscape Science Directorate, Environment Canada, 867 Lakeshore Road, Burlington, Ontario, Canada L7R 4A6

[e]National Institute for Basic Biology, 5-1 Higashiyama, Myodaiji, Okazaki 444-8787, Japan

[f]Rock Hall, Room 284E, University of California, San Francisco, 1550 4th Street, San Francisco, CA 94143-2611, USA

[g]Unit of Research in Organismic Biology, University of Namur (FUNDP), Rue de Bruxelles 61, B-5000 Namur, Belgium

[h]Faculty of Food Technology and Biotechnology University of Zagreb, Pierotti St. 6, Zagreb, Croatia

[i]Department Chemistry & Chemical Engineering, Royal Military College of Canada, P.O. Box 17 000, Stn Forces, Kingston, ON, Canada K7K 7B4

[j]Laboratorio de Endocrinología y Tumores Hormonodependientes, School of Biochemistry and Biological Sciences, Universidad Nacional del Litoral, Santa Fe, Argentina

[k]Integrative Biology Department, University of South Florida, Tampa, FL 33620, USA

[l]Institute of Environmental Ecology, IDEA Consultants, Inc., 1334-5 Riemon, Ooigawa, Shida, Shizuoka 421-0212, Japan

*The article was originally published in the „Journal of Steroid Biochemistry & Molecular Biology“ 127 (2011) 64– 73 by the publisher Elsevier. The article has been approved for scholarly and non-commercial use. http://www.ncbi.nlm.nih.gov/pubmed/21419222 The reprint permission granted by the authors.

[m]Department of Morphology, Federal University of Minas Gerais, Cx. Postal 486, CEP 31.270-901 Belo Horizonte, MG, Brazil

[n]Centre for Toxicology, The School of Pharmacy, 29/39 Brunswick Square, London WC1N 1AX, UK

[o]Department of Biology, Concordia University, Montreal, PQ, Canada H3G 1M8

[p]Department of Medical Oncology, Dana-Faber Cancer Institute and Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

[q]Centre for Advanced Research in Environmental Genomics, Department of Biology, University of Ottawa, 20 Marie Curie, Ottawa, Ontario, Canada K1N 6N5

[r]6811 Cypress Pt N, Austin, TX 78746, USA

Atrazine is the most commonly detected pesticide contaminant of ground water, surface water, and pre- cipitation. Atrazine is also an endocrine disruptor that, among other effects, alters male reproductive tissues when animals are exposed during development. Here, we apply the nine so-called "Hill criteria" (Strength, Consistency, Specificity, Temporality, Biological Gradient, Plausibility, Coherence, Experiment, and Analogy) for establishing cause–effect relationships to examine the evidence for atrazine as an endocrine disruptor that demasculinizes and feminizes the gonads of male vertebrates. We present exper- imental evidence that the effects of atrazine on male development are consistent across all vertebrate classes examined and we present a state of the art summary of the mechanisms by which atrazine acts as an endocrine disruptor to produce these effects.

Atrazine demasculinizes male gonads producing testicular lesions associated with reduced germ cell numbers in teleost fish, amphibians, reptiles, and mammals, and induces partial and/or complete fem- inization in fish, amphibians, and reptiles. These effects are strong (statistically significant), consistent across vertebrate classes, and specific. Reductions in androgen levels and the induction of estrogen syn- thesis – demonstrated in fish, amphibians, reptiles, and mammals – represent plausible and coherent mechanisms that explain these effects. Biological gradients are observed in several of the cited studies, although threshold doses and patterns vary among species. Given that the effects on the male gonads described in all of these experimental studies occurred only after atrazine exposure, temporality is also met here. Thus the case for atrazine as an endocrine disruptor that demasculinizes and feminizes male vertebrates meets all nine of the "Hill criteria".

Article submitted for the special issue on Endocrine diruptors.

Key words:

Atrazine, Gonads, Endocrine disruptor

1. Introduction

Atrazine is a triazine herbicide used primarily on corn [1]. Atrazine is the most commonly detected pesticide contaminant of ground, surface, and drinking water [1–12], and can even be found in rainwater [13–18]. As early as 1997, Crain et al. [19] suggested that atrazine is an endocrine disruptor capable of inducing aromatase and leading to inappropriate and excess estrogen production and in 1998 Reeder et al reported an association between atrazine and intersex gonads in amphibians in the wild [20]. Shortly after this initial report, in 2000, Sanderson et al. [21–23] characterized the effect of atrazine on aromatase in more detail and suggested that "a logical concern would be that exposure of wildlife and humans to triazine herbicides, which are produced and used in large quantities, and are ubiquitous environmental contaminants, may similarly contribute to estrogen-mediated toxicities and inappropriate sexual differentiation" [23]. In this same year, an EPA study concluded that "atrazine tested positive in the pubertal male screen that the Endocrine-Disrupter Screening and Testing Advisory Committee (EDSTAC) is considering as an optional screen for endocrine disrupters" [24].

Several studies in amphibians have suggested that atrazine is associated with feminized males in the wild [25–27]. In field studies, atrazine has repeatedly been associated with the presence of testicular oocytes [20,25–27] as well as feminized secondary sex characteristics in male frogs [28]. As recognized by Sir Austin Bradford Hill in 1965, however, "diseases can have more than one cause" [29] and other endocrine disrupters with mechanisms consistent with demasculinization and feminization of animals have been identified [30–32]. In fact, a retrospective study in amphibians showed that testicular oocytes were detected in museum specimens in Illinois prior to the introduction of atrazine [33], so atrazine may only be responsible for a subset of the recent elevations in such findings. Indeed other environmental contaminants have been shown to feminize amphibians also [34–37]. Because of the complexity of exposures to chemical and other stressors in the field, such eco-epidemiological studies must be coupled with controlled laboratory investigations to ensure reliable attribution of observed changes to specific causes. To establish a cause–effect relationship, Hill described nine "criteria" that should be examined [29]: Experimentation, Consistency, Strength, Specificity, Temporality, Biological Gradient, Plausibility, Coherence, and Analogy. Below, we evaluate the evidence for a cause–effect relationship between atrazine and demasculinization and feminization of male vertebrates using these nine criteria and summarize the many documented mechanisms by which atrazine demasculinizes and feminizes exposed vertebrate males.

1.1 Experimentation

Experimentation was Hill's eighth criteria. We examine this criterion first because it provides the strongest support for atrazine as an endocrine disruptor. In fact, regarding experimentation, Hill wrote: "Occasionally it is possible to appeal to experimental, or semi-experimental, evidence" [29]. "Here the strongest support for the causation hypothesis may be revealed" [29]. Experimental evidence is the strongest (according to Hill) because controlled experiments allow scientists to address the strength of the association, consistency, specificity, temporality, biological gradients, and to examine potential mechanisms for plausibility, coherence, and analogy: the remaining eight of the nine so-called "Hill criteria". Below, we present experimental evidence that shows strong, consistent, specific effects of atrazine on male gonadal development and present a plausible coherent mechanism.

Atrazine is a gonadotoxin in males. Atrazine demasculinizes the gonads of exposed male teleost fish, amphibians, reptiles and mammals. We define "demasculinization" of male gonads as a decrease in male gonadal characteristics including decreases in testicular size, decreases in Sertoli cell number, decreases in sperm production, and decreases in androgen production. Atrazine exposure has been reported to disrupt testicular development, resulting in testicular lesions (loss of testicular tissue) in all vertebrates classes examined except birds. In fish, atrazine causes degeneration of interstitial tissue in the testes, but it has not been reported whether germ cells or Sertoli cells are the also targets [38] (Fig. 1). In amphibians [39], reptiles [40] and mammals [41,42], however, atrazine has nearly identical (specific) effects (Fig. 1). Atrazine exposure in these vertebrate classes results in increases in the size of testicular tubules, loss of Sertoli cells, and a marked loss of germ cells, often leaving the testicular tubules empty or only with what appears to be cellular debris ([39–42], Fig. 1).

In addition to the demasculinization effects described above (testicular lesions), atrazine feminizes the gonads of developing male teleost fish (hereafter referred to simply as "fish") [43] (Fig. 2), amphibians [26,44–46], and reptiles [47]. "Feminization" of male gonads is defined as the development of oocytes in the testes or complete ovarian differentiation in genetic males leading to decreases in the frequency of morphologic males in the exposed population. The loss of male germ cells, described above, is accompanied by the development of female germ cells (testicular oocytes) in the testes in some cases ([25,26,43,45,47], Fig. 2). This effect has been reported in fish [43], amphibians [26,44,45], and reptiles [47], but has not been reported in birds or mammals.

Atrazine can also result in complete feminization of males. At least one study in fish [48], three studies in amphibians [39,49,50], and two studies in reptiles [47,51]

have causes significant shifts in sex ratios toward females (Fig. 3). In fish, the effect is manifested by skewed sex ratios in zebrafish (Danio rerio), which have no distinguishable sex chromosomes [48]. In amphibians, shifts in the sex ratio have been documented [49,50], and one study used genetic markers in Xenopus laevis to show that indeed genetic males were converted into complete functional females after exposure to atrazine [39]. Finally, in reptiles, two studies in two different species (of different genera and families) with temperature-dependent sex determination, examined the effect of atrazine on sex differentiation. Atrazine caused female biased sex ratios compared to controls near the transitional male–female temperatures [51], and turtles with testicular oocytes were found only in atrazine treatments [47]. Thus, experimental data support the hypothesis that atrazine both demasculinizes male gonads in all vertebrate classes examined with the possible exception of birds and feminizes the gonads of male vertebrate ectotherms (fish, amphibians, and reptiles).

1.2 Consistency

With regards to consistency (Hill's second criterion), Hill wrote, "Whether chance is the explanation or whether a true hazard has been revealed may sometimes be answered only by a repetition of the circumstances and observations" [29]. Hill queried further, "Has it (the effect) been repeatedly observed by different persons, in different places, circumstances and times?" Moreover, he stated, "I would myself put a good deal of weight upon similar results reached in quite different ways." In fact, this scenario is revealed here: demasculinizing effects of atrazine on developing male gonads across vertebrate classes (fish, amphibians, reptiles, and mammals), as well as partial and complete feminization of male gonads across three vertebrate classes (fish, amphibians, and reptiles). These studies used different routes of exposure, varying doses, and widely varying experimental conditions (see references cited in Figs. 1–3), yet all found similar effects.

Echoing Hill's recommendations, Glen Fox wrote: "In ecoepidemiology, the occurrence of an association in more than one species and species population is very strong evidence for causation" [52]. Kniewald et al. [42] and Victor-Costa et al. [41] independently reported identical effects of atrazine on testes in both Fischer [42] and Wistar [41] rats and, indeed, testicular lesions and feminization of male gonads have been consistently observed across vertebrate classes. These effects on the gonads are both specific and consistent and do not occur merely across populations, species, or even genera or orders, but across vertebrate classes: although the loss of testicular tissue may occur primarily in the interstitium in fish, the loss of Sertoli cells and male germ cells in testicular tubules, development of female germ cells (testicular oocytes) in exposed males and complete feminization of males across three classes of vertebrates are specific and consistent effects described by

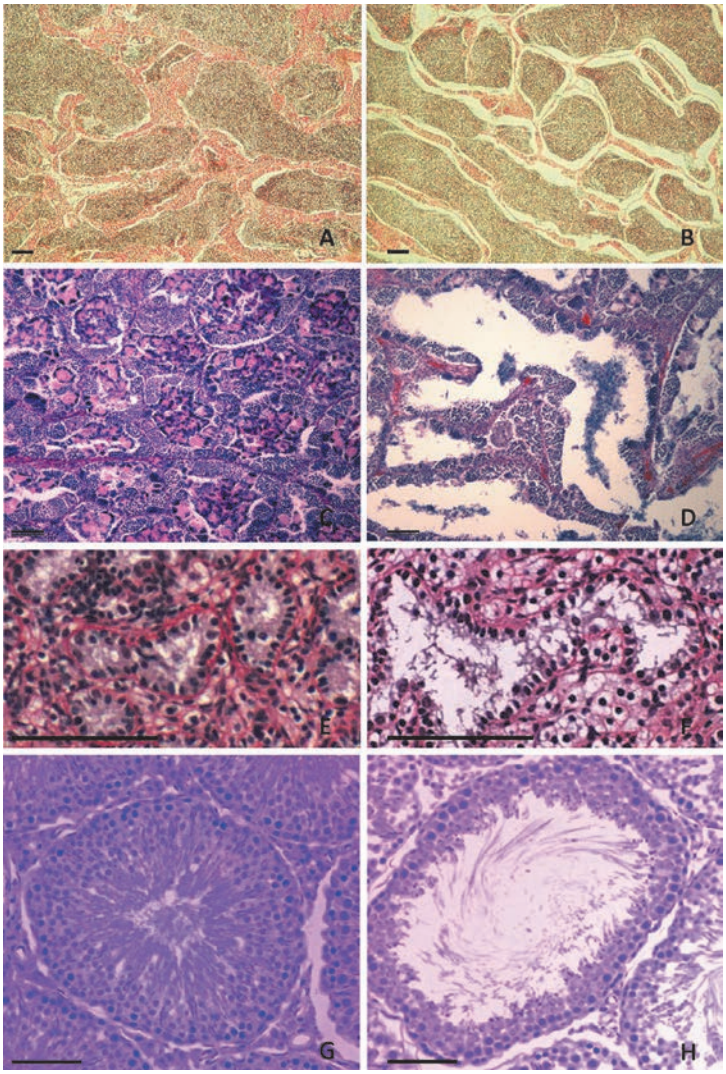


Fig. 1 – Atrazine-induced histologic lesions in testes of vertebrates. Histologic sections from a fish (A and B), an amphibian (C and D), a reptile (E and F) and a mammal (G and H) are shown. Histologic sections of testes of goldfish (Carassius auratus) controls (A) and after 21 days of exposure to water containing atrazine at 1,000 g/L (B). Note the progressive increase in gaps in the interstitium between lobules. Sections were stained with Regaud's hematoxylin, phloxine and light green. For details see Spano et al. [38]. Histological section of testes in African clawed frogs (Xenopus laevis) controls (C) and after exposure to atrazine at 2.5 g/L throughout larval and postmetamorphic development (D). Sections were stained in Harris' hematoxylin and eosin. For details see Hayes et al. [39]. Photomicrographs showing seminiferous tubules from a vehicle (E) and an atrazine-treated (F) caiman. Tissue sections were stained with Picrosirius solution and counterstained with Harris hematoxylin. For details, see Rey et al. [40]. Testicular tubules of control rats (G) and tubule of rats given atrazine at 200 mg/kg by gavage for 15 days (H). Atrazine-exposed rats were characterized by luminal dilation. Extended dosing up to 40 days resulted in testicular atrophy, which was mostly formed by Sertoli-only tubules (not shown). Sections were stained with hematoxylin and eosin. For details see Victor-Costa et al. [41]. Bar = 100 in all panels.

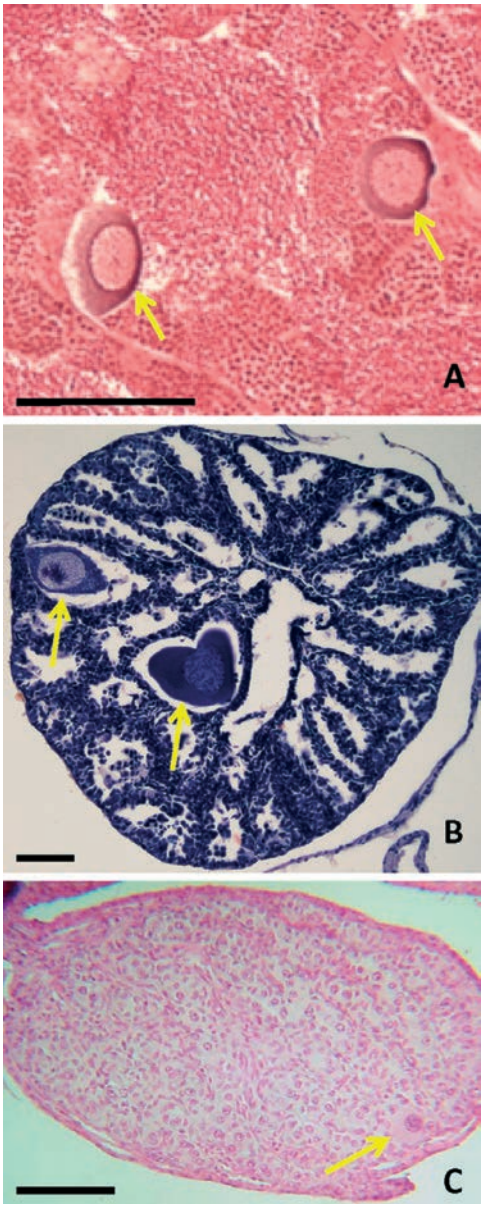


Fig. 2 – Partial feminization by atrazine in vertebrates. Testicular oocytes are induced by atrazine in fish (A), amphibians (B), and reptiles (C). Testes from an adult fathead minnow (Pimephales promelas) (A) exposed to water containing atrazine at 5 g/L for 14 days presenting multiple testicular oocytes within the gametogenic and support- ive cellular structures of the testes (Photo courtesy Diana Papoulias, U.S. Geological Survey, Columbia Environmental Research Center, Columbia, MO, USA). See Tillitt et al. [43] for details of experimental conditions. Testicular oocytes in the testes of a male Rana pipiens [B] exposed to atrazine at 0.1 g/L. Section stained in Harris' hematoxylin and eosin. For details see Hayes et al. [26]. Testicular oocyte (stained in Harris' hematoxylin and eosin) in a snapping turtle (Chelydra serpentine) exposed to soil treated with atrazine at a typical application rate (3.1 L/ha). For details see de Solla et al. [47]. Bar = 100 in all panels.

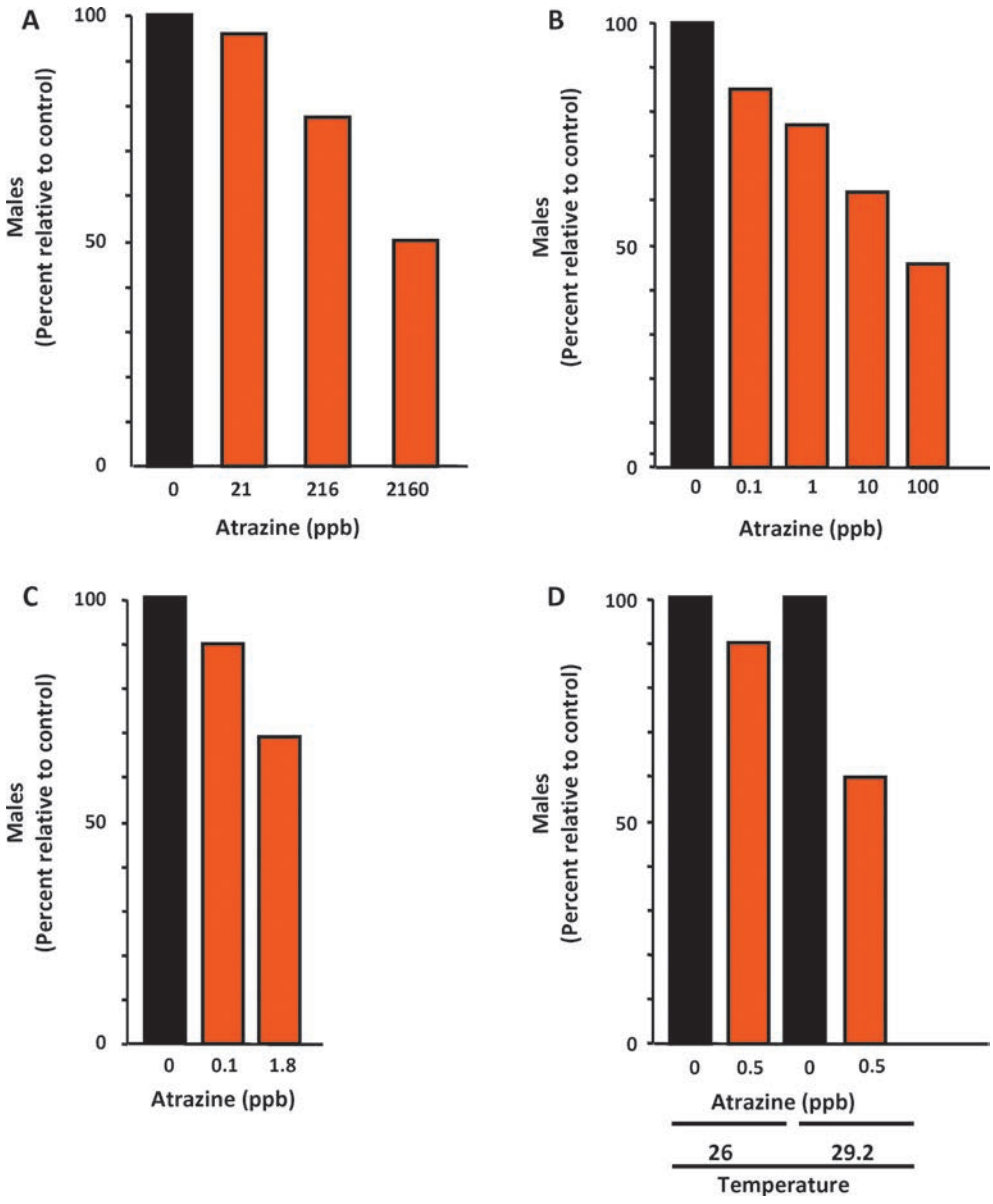


Fig. 3 – Complete sex reversal by atrazine in vertebrates. Atrazine exposure causes a loss of males in exposed fish (A), amphibians (B and C), and reptiles (D). In fish, atrazine exposure produced a concentration-dependent decrease in the frequency of males [48]. (A) Similarly atrazine produced a concentration-dependent decrease in the frequency of males in African clawed frogs (Xenopus laevis) in the laboratory [49] (B) and in leopard frogs (Rana pipiens) exposed in semi-field conditions [50] (C). In red-eared sliders (Trachemys scripta elegans), atrazine seemed to reduce the number of males slightly at 26° , and a statistically significant reduction was evident at 29.2° [51] (D). In all cases, data are shown as percent males in experimental groups relative to controls. Figures A and B were adapted from Hayes et al. [39]. Original data for panel C from Langlois et al. [50] and original data were from panel D from Willingham et al. [51].

independent laboratories in eight different countries on five continents (see references cited in Figs. 1–3 and work described herein). These effects are also consistent with earlier studies in African clawed frogs (Xenopus laevis) which showed a loss of germ cells and nursing (Sertoli) cells in both males and females exposed during larval stages [53,54].

That is not to say, however, that these effects have been documented in all species in these classes or even in all populations (or studies) within a species. Variation is to be expected and as Hill pointed out: "the different results of a different inquiry certainly cannot be held to refute the original evidence." [29]. To quote Hill, "The lesson here is that broadly the same answer has been reached in quite a wide variety of situations and techniques. In other words, we can justifiably infer that the association is not due to some constant error or fallacy that permeates every inquiry." [29]

1.3 Temporality, Specificity, and Strength

The effects described here meet all three of these components of Hill's criteria. Regarding temporality, Hill believed that the "cause" should precede the effect. In the current case, atrazine exposure should precede demasculinization and feminization. In the experimental evidence presented here, this of course is the case.

In addition, the lesions of the gonads (demasculinization), partial feminization (testicular oogenesis) and complete feminization (sex reversal resulting in female-biased sex ratios) are specific effects. It is important to realize, as Hill pointed out, that diseases may have more than one cause, and a given factor (such as atrazine) may produce more than one disease (effect). This observation is certainly true for atrazine. More than one chemical may induce the gonadal malformations that atrazine induces (see references in Section 1), and atrazine affects more than gonadal development (see Section 2). In the cases presented here, however, atrazine's effects are supported by controlled experiments in fish, amphibians, reptiles, and mammals, and produce similar specific effects across studies. Thus, the complications associated with identifying cause and effect in epidemiological and eco-epidemiological studies are not of general concern here.

The effects are also strong associations, in the species examined. Hill listed "strength" as his first criterion. Despite the importance of strength in establishing cause and effect, Hill wrote: "In thus putting emphasis upon the strength of an association, we must, nevertheless look at the obverse of the coin. We must not be too ready to dismiss a cause–effect hypothesis merely on the grounds that the observed association appears to be slight" [29]. Hill later goes on, regarding statistics, to caution, ". . .far too often we deduce 'no difference' from 'no significant differ-

ence"'. In the cases presented here, all of the effects are statistically significant, with the exception of some of the findings in snapping turtles (Chelydra serpentina) [47]. But as Hill espoused, statistics are not even necessary in obvious cases, such as here, where the malformations described do not occur in controls in any of the experiments. In the few studies where malformed gonads were observed in controls, the authors reported atrazine contamination in controls above the biologically effective dose [55–57], with the exception of Orton et al. who found intersex animals in control Rana pipiens [46].

1.4 Biological gradient

Here, Hill suggested that some dose–response relationship would lend support for the cause–effect relationship. Although Hill focused on a monotonic linear dose response, he readily acknowledged that in some cases "a much more complex relationship to satisfy the cause and effect hypothesis" must be envisaged. With regard to feminization of males, several studies show just the type of dose response that Hill suggested would support cause and effect. In both zebrafish and African clawed frogs, the frequency of males declines in a dose-dependent fashion in response to increasing doses of atrazine (Fig. 3). In most studies, the proportion of affected males increases with atrazine concentration [26,44–46,56]. With regard to demasculinization (testicular lesions) and partial feminization (testicular oogenesis), quantitative methods are yet to be standardized, thus parametric dose–response analyses of this type are not available.

Regarding the effective doses, the demasculinization effects of atrazine were produced at low ecologically relevant doses (e.g. 2.5 ppb or below) in amphibians. Doses in reptiles are more difficult to interpret because the doses reported in some of these studies are the doses applied to the egg shell and it is not known how much atrazine reached the affected tissues. Nonetheless, snapping turtle eggs readily absorb current-use pesticides, including atrazine, from soil treated with typical application rates (de Solla et al., unpublished data). We hypothesize, however, that the barrier created by the egg shell and membranes may explain why effects in reptiles are less pronounced. In addition, it is possible that the exposure time to atrazine was insufficient in some reptile studies. In rats, the atrazine was delivered by gavage and the higher metabolism [58] in small endotherms likely explains why seemingly higher doses are required for effects in rodents, but comparative studies of atrazine uptake, distribution in the body, and metabolism are not available.

1.5 Plausibility and Coherence

Plausible, coherent mechanisms are available to explain gonadal demasculinization and feminization. Atrazine exposure significantly reduces synthesis, secretion and circulating levels of androgens across vertebrate classes including fish [38,59], amphibians [26,39], reptiles [40], and mammals [24,60] with lesser effects in birds [61]. Androgen production is critical for germ cell differentiation, development and maturation. Thus, limiting androgen production and availability provides a plausible mechanism to explain the demasculinization of gonads in exposed males. Multiple mechanisms likely account for the decreased androgens and decreased androgen activity (summarized in Fig. 4): (1) atrazine inhibits luteinizing hormone (LH) and follicle stimulating hormone (FSH) peaks and surges by inhibiting pulsatile release of gonadotropin releasing hormone (GnRH) from the hypothalamus which leads to decreased androgen synthesis [24,62–66]; (2) atrazine inhibits LH release from the pituitary directly which leads to decreased androgen production [62,64,67,68]; (3) atrazine inhibits the enzyme 5 reductase which leads to decreased levels of the potent androgen dihydrotestosterone (DHT) and leaves more testoster-

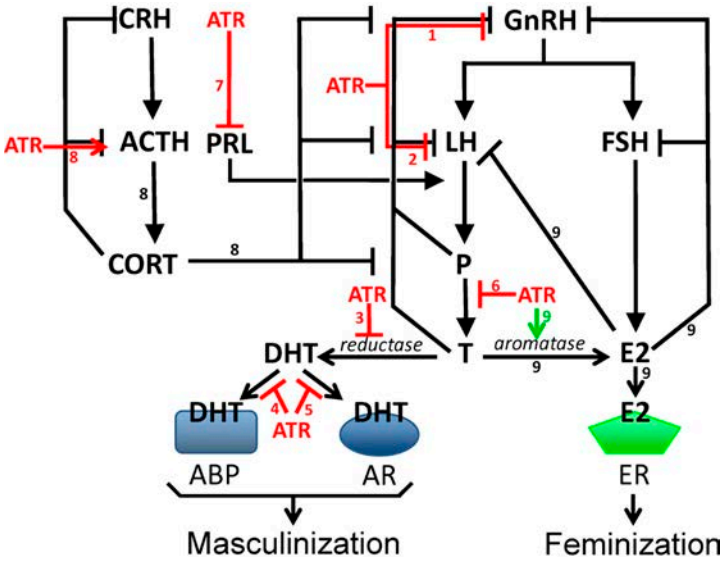


Fig. 4 – Multiple mechanisms of action have been identified for atrazine's demasculinizing and feminizing effects on male gonads. Arrows indicate processes that are increased; bars indicate processes that are inhibited. Red lines indicate demasculinizing pathways that are directly affected by ATR and green lines indicate feminizing pathways that are directly affected by ATR. Numbers on pathways, refer to mechanisms listed in the text (see Section 1.5). ABP = androgen binding protein, ACTH = adrenocorticotropic hormone, AR = androgen receptor, ATR = atrazine, CORT = cortisol/corticosterone, CRH = corticotrophin-releasing hormone, DHT = dihydrotestosterone, E2 = 17 estradiol, ER = estrogen receptor, FSH = follicle stimulating hormone, GnRH = gonadotropin stimulating hormone, LH = luteinizing hormone, P = progesterone, PRL = prolactin., and T = testosterone (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

one as substrate for aromatase to convert to estradiol which negatively feeds back on the hypothalamus and pituitary [69,70]; (4) atrazine inhibits binding of DHT to the androgen binding protein [71]; (5) atrazine inhibits interactions between DHT and the androgen receptor(AR) [69,70,72] but perhaps not by directly binding to the receptor [73], but perhaps not by directly inhibiting binding to the receptor (see [78]); (6) atrazine inhibits androgen synthesis in the testes [60,74,75]; (7) atrazine decreases prolactin secretion [63,78,79]. Prolactin promotes LH receptor expression, and thus a decrease in prolactin would lead to a decrease in LH receptors, impairing normal LH-stimulation of testosterone production; (8) atrazine increases adreno-corticotropic hormone (ACTH) secretion from the pituitary leading to increased progesterone and increased corticosteroid (cortisol or corticosterone) secretion [76,77]. Progesterone negatively feeds back on GnRH, LH and FSH and thus decreases androgen production and reproductive function, whereas corticosteroids inhibit the reproductive axes at the hypothalamo (GnRH), anterior pituitary (LH and FSH) and the gonads (androgen production).

The partial and complete feminization of gonads in fish, amphibians and reptiles are analogous to the effects of estrogen in these vertebrate classes (see Section 1.6). Atrazine could feminize animals by increasing estrogen synthesis, decreasing degradation of endogenous estrogen, or acting as an estrogen receptor agonist. No available evidence suggests that atrazine decreases degradation of endogenous estrogen and this is not a plausible mechanism anyway, because developing males would not have adequate circulating estrogens to stabilize. In addition, atrazine does not bind the estrogen receptor [23,80], so this is not a plausible mechanism either. Several studies, however, suggest that atrazine increases estrogen synthesis. Atrazine induces aromatase in the gonads of fish [48], amphibians [39], and reptiles (in vitro) [19], and in human cell lines [21–23,48,81–83]. Also, atrazine increases circulating estrogens in fish [38,59], amphibians [39], and in mammals (laboratory rats) [24]. Estrogens induce partial and complete feminization in fish [84,85], amphibians [86,87], and reptiles [86,88], so the induction of aromatase and subsequent increases in estrogen synthesis represent a plausible mechanism for the feminization effects.

In addition, atrazine and the atrazine metabolite, deethylatrazine, also inhibit 5α-reductase [66]. Reducing the availability of 5α-reductase in atrazine-exposed male rats, decreases the conversion of testosterone to the potent androgen 5α-dihydrotestosterone [72]. Dihydrotestosterone is not convertible to estrogen; thus, the inhibition of 5α-reductase by atrazine may increase the pool of testosterone available for conversion to estrogen. For example exposure to exogenous testosterone, though not dihydrotestosterone, can cause feminization in turtles [89], through the conversion to estrogen via aromatase. Given the similarity between the effects of exogenous estrogen in vertebrates (see Section 1.6) and atrazine, this plausible mechanism (the induction of aromatase) is coherent.

1.6 Analogy

Hill wrote, ". . .the cause-and-effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease". Androgens are necessary for testicular development and maintenance of male germ cells in all vertebrates. Thus, given that atrazine reduces androgen production and stability, it is reasonable to expect the demasculinization effect in all vertebrates. On the other hand, partial or complete sex reversal of gonads by estrogens is limited to fish and amphibians, which lack morphological distinguishable sex chromosomes, and to reptiles with environmental sex determination, which lack sex chromosomes altogether [86]. Birds and mammals, which have genetic sex determination and highly differentiated sex chromosomes are not susceptible to estrogen-induced sex reversal of the gonads [86,90]. As such, while depleting androgens will impair testicular development and induce testicular lesions (such as the effects described here) in all vertebrates, increasing estrogen production (via atrazine) would not be expected to induce feminization of the gonads in birds and mammals, but would do so in fish, amphibians, and reptiles with environmental sex determination.

2. Conclusions

All nine of the Hill criteria are met in the case for atrazine as an endocrine disruptor that alters male gonadal development. Furthermore, effects occur in all vertebrate classes examined with the possible exception of birds, suggesting a ubiquitous problem. Published studies assessing the effects of atrazine on the developing gonads of birds are limited to one study in Japenese quail (Coturnix coturnix japonica) [91] and another in chickens (Gallus gallus domesticus) [92]. When injected into quail eggs at 123, 246, and 504 μg/kg, atrazine had no effect on the sex ratio as determined on day 14 post-hatching, but 504 μg/kg caused a decrease in ovarian weight and in progesterone levels in females [91]. Atrazine had no effect on gonadal weight or circulating estradiol, testosterone, or progesterone in males and no incidences of left ovotestes were observed [91]. In contrast, when atrazine was injected into chicken eggs [92], 0.1 to 3 mg/egg caused retention of the right gonad in female chicks [92]. No effects were observed in males [92]. So based on these two studies, atrazine produces much more subtle effects in birds, relative to other vertebrate taxa and only at high doses. Birds are also probably less likely to be exposed to atrazine in the wild. Given that atrazine is mostly an aqueous contaminant, water fowl maybe more likely affected by atrazine and studies are warranted. Similarly, the absence of data in the only two remaining classes of vertebrates

(Agnatha and Chondricthyes) reflects the fact that no published studies on effects of atrazine are available for these taxa.

While morphologic changes found in studies of atrazineexposed vertebrates are of concern, functional impairments are likely of greatest importance. In fact, male salmon (Salmo salar) exposed to atrazine showed decreased androgen levels, decreased mating behavior, and decreased milt production [59]. Nearly identical reproductive impairments were observed in amphibians (decreased androgens, decreased mating behavior and decreased fertility) [39]. Similarly, atrazine induced poor reproductive performance in rodents [42] and resulted in as much as a 50% decrease in epididymal sperm number and decreased sperm motility [42]. Low fertility, low sperm count, and poor semen quality were also associated with atrazine exposure (as measured by atrazine metabolites in the urine) in humans living in agricultural areas where atrazine was widely used [93]. Furthermore, atrazine is also associated with follicular atresia in females of two species of fish [38,43], limiting the reproductive potential of females as well.

Atrazine is prevalent and persistent in the environment. There are many other documented reproductive effects of atrazine in laboratory rodents: induced abortion [62,64], impaired mammary development [94–96], the induction of reproductive and hormonedependent cancers [97–102] as well as other non-reproductive effects including impaired immune function (also observed in multiple studies across vertebrate classes) [103–127] and impaired neural development [117,128–132]. Thus, with the additional the indirect effects of atrazine on habitats [133–142], atrazine can have dramatic effects on ecosystems, environmental health and public health.

Acknowledgements

We thank Donald Tillitt and Diana Papoulias, USGS, for the photograph featured in Fig. 2, panel A. Primary research on atrazine was supported as follows: TBH (Novartis, Syngenta Crop Protection, Ecorisk, the National Science Foundation, the World Wildlife Fund, the Alton Jones Foundation, the Homeland Foundation, the Rose Foundation, Park-Water Company, the Biofaculty Award [UCB], the Distinguished Mentor Award [UCB], the Distinguished Teaching Award [UCB], the Mitchell Kapor Foundation, the David Foundation, the Cornell-Douglas Foundation, the Wallace Global Fund, the Class of ‘43 Endowed Chair (UCB), the Toxic Substances Research and Teaching Program (UCB), Hewlett Packard, the Biology Fellows Program (Howard Hughes Medical Institute), the McNair Scholars Program (UCB), the Amgen Scholars Program (UCB), and the Mentored Research fellowship program (UCB); LA (USDA grant through the Center for Designing Foods to Improve Nutrition; VRB (John G. Shedd Aquarium Funding through support from the Dr. Scholl Foundation); SRdS (Environment Canada); TI & TO

(Ministry of the Environment, Japan); HI & MS (National Institutes of Health), PK (National Science Funds, FNRS, Belgium); JK & ZK (Ministry of Science and Technology of the Republic of Croatia, Former Yu-USA Joint Fund for S&T Coop, and the Ford Foundation; VSL (NSERC-PGSD program), EHL & MMT (Universidad Nacional del Litoral, CAI+D program, the Argentine National Council for Science and Technology (CONICET), and the Argentine National Agency for the Promotion of Science and Technology; KAM (National Institute of Environmental Health Sciences to L. Guillette, Rewald, Olowo, Society for Integrative and Comparative Biology, Sigma Xi grants in aid of research and the University of Florida Institute of Food and Agricultural Sciences Women's Club Fellowship; CO & ABV-C (Conselho Nacional de Desenvolvimento Cientifico e Tecnológico, CNPq/Brazil, grant and research fellowship to CAO, Master fellowship to ABVC supported by Coordenac¸ão de Aperfeic¸oamento de Pessoal de Nível Superior, CAPES/Brazil); FO (Declining Amphibian Population Task Force and Depertment for Environmental, Food and Rural Affairs, UK); SR & LET-M (Toxic Substances Research Initiative, Government of Canada; VLT (Environment Canada and the Canadian Water Network); EW (Texas State University, San Marcos).

References

[1] K. Solomon, et al., Ecological risk assessment of atrazine in North American surface waters, Environ. Toxicol. Chem. 15 (1996) 31–76.

[2] R. Boyd, Herbicides and herbicide degradates in shallow groundwater and the Cedar River near a municipal well field, Cedar Rapids, Iowa, Sci. Total Environ. 248 (2000) 241–253.

[3] P. Capel, S. Larson, Effect of scale on the behavior of atrazine in surface waters, Environ. Sci. Technol. 35 (4) (2001) 648–657.

[4] L.H. Du Preez, et al., Seasonal exposures to triazine and other pesticides in surface waters in the western Highveld corn-production region in South Africa, Environ. Pollut. 135 (1) (2005) 131–141.

[5] J. Fenelon, R. Moore, Transport of agrichemicals to ground and surface waters in a small central Indiana watershed, J. Environ. Qual. 27 (1998) 884–894.

[6] J. Fischer, B. Apedaile, L. Vanclief, Seasonal loadings of atrazine and metolachlor to a southeastern Ontario river from surface runoff and groundwater discharge, Water Qual. Res. J. Canada 30 (3) (1995) 533–553.

[7] R. Frank, L. Logan, Pesticide and industrial chemical residues at the mouth of the Grand, Saugeen and Thames rivers, Ontario, Canada, 1981–85, Arch. Environ. Contam. Toxicol. 17 (1988) 741–754.

[8] R. Frank, L. Logan, B. Clegg, Pesticide and polychlorinated biphenyl residues in waters at the mouth of the Grand, Saugeen and Thames rivers, Ontario, Canada, 1986–1990, Arch. Environ. Contam. Toxicol. 21 (1991) 585–595.

[9] R. Frank, et al., Survey of farm wells for pesticides, Ontario, Canada, Arch. Environ. Contam. Toxicol. 16 (1987) 1–8.

[10] R. Frank, G. Sirons, Atrazine: its use in corn production and its loss to stream waters in southern Ontario, Sci. Total Environ. 12 (1979) 223–239.

[11] M. Graymore, F. Stagnitti, G. Allinson, Impacts of atrazine in aquatic ecosystems, Environ. Int. 26 (7–8) (2001) 483–495.

[12] D. Rudolph, M. Goss, Ontario farm groundwater quality survey – summer 1992, A report prepared for Agriculture Canada, Agri-Food Development Branch, Guelph, Ontario, Agriculture Canada, 1993.

[13] J.L. Hatfield, et al., Herbicide and nitrate distribution in central Iowa rainfall, J. Environ. Qual. 25 (2) (1996) 259–264.

[14] O. Lode, et al., Pesticides in precipitation in Norway, Sci. Total Environ. 160–161 (1995) 421–431.

[15] M.A. Mast, W.T. Foreman, S.V. Skaates, Current-use pesticides and organochlorine compounds in precipitation and lake sediment from two high-elevation national parks in the Western United States, Arch. Environ. Contam. Toxicol. 52 (3) (2007) 294–305.

[16] S. Miller, et al., Atrazine and nutrients in precipitation: results from the Lake Michigan mass balance study, Environ. Sci. Technol. 34 (1) (2000) 55–61.

[17] J.R. Vogel, M.S. Majewski, P.D. Capel, Pesticides in rain in four agricultural watersheds in the United States, J. Environ. Qual. 37 (3) (2008) 1101–1115.

[18] E. Thurman, A. Cromwell, Atmospheric transport, deposition, and fate of triazine herbicides and their metabolites in pristine areas at Isle Royale National Park, Environ. Sci. Technol. 34 (2000) 3079–3085.

[19] D. Crain, et al., Alterations in steroidogenesis in alligators (*Alligator mississippiensis*) exposed naturally and experimentally to environmental contaminants, Environ. Health Perspect. 105 (1997) 528–533.

[20] A. Reeder, et al., Forms and prevalence of intersexuality and effects of environmental contaminants on sexuality in cricket frogs (Acris crepitans), Environ. Health Perspect. 106 (1998) 261–266.

[21] J. Sanderson, et al., Induction and inhibition of aromatase (CYP19) activity by various classes of pesticides in H295R human adrenocortical carcinoma cells, Toxicol. Appl. Pharmacol. 182 (2002) 44–54.

[22] J.T. Sanderson, et al., Effects of chloro-s-triazine herbicides and metabolites on aromatase activity in various human cell lines and on vitellogenin production in male carp hepatocytes, Environ. Health Perspect. 109 (2001) 1027– 1031.

[23] J.T. Sanderson, et al., 2-Chloro-triazine herbicides induce aromatase (CYP19) activity in H295R human adrenocortical carcinoma cells: a novel mechanism for estrogenicity? Toxicol Sci. 54 (2000) 121–127.

[24] T. Stoker, et al., The effect of atrazine on puberty in male Wistar rats: an evaluation in the protocol for the assessment of pubertal development and thyroid function, Toxicol. Sci. 58 (1) (2000) 50–59.

[25] T.B. Hayes, et al., Feminization of male frogs in the wild, Nature 419 (2002) 895–896.

[26] T.B. Hayes, et al., Atrazine-induced hermaphroditism at 0.1 ppb in American leopard frogs (Rana pipiens): laboratory and field evidence, Environ. Health Perspect. 111 (2002) 568–575.

[27] M.B. Murphy, et al., Atrazine concentrations, gonadal gross morphology and histology in ranid frogs collected in Michigan agricultural areas, Aquat. Toxicol. 76 (3–4) (2006) 230–245.

[28] K.A. McCoy, et al., Agriculture alters gonadal form and function in the toad Bufo marinus, Environ. Health Perspect. 116 (11) (2008) 1526–1532.

[29] A. Hill, The environment and disease: association or causation, Proc. R. Soc. Med. 58 (1965) 295–300.

[30] IPCS global assessment of the state-of-the-science of endocrine disruptors, in: International Program of Chemical Safety, World Health Organization, Geneva, 2002.

[31] A Daxenberger, Pollutants with androgen-disrupting potency, Eur. J. Lipid Sci. Technol. 104 (2) (2002) 124–130.

[32] C.M. Markey, et al., Endocrine disruptors: from Wingspread to environmental developmental biology, J. Steroid Biochem. Mol. Biol. 83 (1–5) (2002) 235–244.

[33] A.L. Reeder, et al., Intersexuality and the cricket frog decline: historic and geographic trends, Environ. Health Perspect. 113 (3) (2005) 261–265.

[34] A.D. Sowers, M.A. Mills, S.J. Klaine, The developmental effects of a municipal wastewater effluent on the northern leopard frog, Rana pipiens. Aquatic Toxicology 94 (2) (2009) 145–152.

[35] C. Howe, et al., Toxicity of glyphosate-based pesticides to four North American frog species, Environ. Toxicol. Chem. 23 (8) (2004) 1928–1938.

[36] N.S. Hogan, et al., Estrogenic exposure affects metamorphosis and alters sex ratios in the northern leopard frog (*Rana pipiens*): Identifying critically vulnerable periods of development, General and Comparative Endocrinology 156 (3) (2008) 515–523.

[37] M.B. Jofre, W.H. Karasov, Effect of mono-ortho and di-ortho substituted polychlorinated biphenyl (PCB) congeners on leopard frog survival and sexual development, Chemosphere 70 (9) (2008) 1609–1619.

[38] L. Spano, et al., Effects of atrazine on sex steroid dynamics, plasma vitellogenin concentration and gonad development in adult goldfish (*Carassius auratus*), Aquat. Toxicol. (Amsterdam) 66 (4) (2004) 369–379.

[39] T.B. Hayes, et al., Atrazine induces complete feminization and chemical castration in male African clawed frogs (*Xenopus laevis*), Proc. Natl. Acad. Sci. U.S.A. 107 (10) (2010) 4612–4617.

[40] F. Rey, et al., Prenatal exposure to pesticides disrupts testicular histoarchitecture and alters testosterone levels in male Caiman latirostris, Gen. Comp. Endocrinol. 162 (3) (2009) 286–292.

[41] A.B. Victor-Costa, et al., Changes in testicular morphology and steroidogenesis in adult rats exposed to Atrazine, Reprod. Toxicol. 29 (3) (2010) 323–331.

[42] J. Kniewald, et al., Disorders of male rat reproductive tract under the influence of atrazine, J. Appl. Toxicol. 20 (2000) 61–68.

[43] D.E. Tillitt, et al., Atrazine reduces reproduction in fathead minnow, Marine Environ. Res. 66 (1) (2008) 51–151.

[44] T.B. Hayes, et al., Hermaphroditic, demasculinized frogs after exposure to the herbicide atrazine at low ecologically relevant doses, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 5476–5480.

[45] T.B. Hayes, et al., Characterization of atrazine-induced gonadal malformations and effects of an androgen antagonist (*cyproterone acetate*) and exogenous estrogen (estradiol 17): support for the demasculinization/feminization hypothesis, Environ. Health Perspect. 114 (Suppl. 1) (2006) 134–141.

[46] F. Orton, J. Carr, R. Handy, Effects of nitrate and atrazine on larval development and sexual differentiation in the northern leopard frog Rana pipiens, Environ. Toxicol. Chem. 25 (1) (2006) 65–71.

[47] S.R. De Solla, et al., Effects of environmentally relevant concentrations of atrazine on gonadal development of snapping turtles (*Chelydra serpentina*), Environ. Toxicol. Chem. 25 (2) (2006) 520–526.

[48] M. Suzawa, H. Ingraham, The herbicide atrazine activates endocrine gene networks via non-steroidal NR5A nuclear receptors in fish and mammalian cells, PLoS One 3 (2008) 2117.

[49] T. Oka, et al., Effect of atrazine on metamorphosis and sexual differentiation in Xenopus laevis, Aquat. Toxicol. 87 (4) (2008) 215–226.

[50] V. Langlois, et al., Low levels of the herbicide atrazine alters sex ratios and reduces metamorphic success in Rana pipiens tadpoles raised in outdoor mesocosms, Environ. Health Perspect. (2009).

[51] E.J. Willingham, The effects of atrazine and temperature on turtle hatchling size and sex ratios, Front. Ecol. Environ. 3 (6) (2005) 309–313.

[52] G. Fox, Practical causal inference for epidemiologists, J. Toxicol. Environ. Health 33 (1991) 359–373.

[53] L. Tavera-Mendoza, et al., Response of the amphibian tadpole (*Xenopus laevis*) to atrazine during sexual differentiation of the testis, Environ. Toxicol. Chem. 21 (2002) 527–531.

[54] L. Tavera-Mendoza, et al., Response of the amphibian tadpole *Xenopus laevis* to atrazine during sexual differentiation of the ovary, Environ. Toxicol. Chem. 21 (6) (2002) 1264–1267.

[55] T.B. Hayes, There is no denying this: defusing the confusion about atrazine, Bioscience 54 (12) (2004) 1138–1149.

[56] J. Carr, et al., Response of larval *Xenopus laevis* to atrazine: assessment of growth, metamorphosis, and gonadal and laryngeal morphology, Environ. Toxicol. Chem. 22 (2003) 396–405.

[57] M. Hecker, et al., Plasma concentrations of estradiol and testosterone, gonadal aromatase activity and ultrastructure of the testis in *Xenopus laevis* exposed to estradiol or atrazine, Aquat. Toxicol. (Amsterdam) 72 (4) (2005) 383–396.

[58] T. Gojmerac, J. Kniewald, Atrazine biodegradation in rats – a model for mammalian metabolism, Bull. Environ. Contam. Toxicol. 43 (2) (1989) 199–206.

[59] A. Moore, C. Waring, Mechanistic effects of a triazine pesticide on reproductive endocrine function in mature male Atlantic salmon (*Salmo salar L.*) parr., Pesticide Biochem. Physiol. 62 (1998) 41–50.

[60] A. Friedmann, Atrazine inhibition of testosterone production in rat males following peripubertal exposure, Reprod. Toxicol. 16 (3) (2002) 275–279.

[61] K.W. Wilhelms, et al., Effects of atrazine on sexual maturation in female Japanese quail induced by photostimulation or exogenous gonadotropin, Environ. Toxicol. Chem. 25 (1) (2006) 233–240.

[62] A. Cummings, B. Rhodes, R. Cooper, Effect of atrazine on implantation and early pregnancy in 4 strains of rats, Toxicol. Sci. 58 (1) (2000) 135–143.

[63] R.L. Cooper, et al., Atrazine disrupts the hypothalamic control of pituitary–ovarian function, Toxicol. Sci. 53 (2000) 297–307.

[64] M. Narotsky, et al., Strain comparisons of atrazine-induced pregnancy loss in the rat, Reprod. Toxicol. 15 (1) (2001) 61–69.

[65] C.D. Foradori, et al., Effects of atrazine and its withdrawal on gonadotropinreleasing hormone neuroendocrine function in the adult female Wistar rat, Biol. Reprod. 81 (6) (2009) 1099–1105.

[66] T. Babic-Gojmerac, Z. Kniewald, J. Kniewald, Testosterone metabolism in neuroendocrine organs in male rats under atrazine and deethylatrazine influence, J. steroid. Biochem. 33 (1) (1989) 141–146.

[67] C.D. Foradori, et al., Atrazine inhibits pulsatile luteinizing hormone release without altering pituitary sensitivity to a gonadotropin-releasing hormone receptor agonist in female Wistar rats, Biol. Reprod. 81 (1) (2009) 40–45.

[68] T. McMullin, et al., Evidence that atrazine and diaminochlorotriazine inhibit the estrogen/progesterone induced surge of luteinizing hormone in female Sprague–Dawley rats without changing estrogen receptor action, Toxicol. Sci. 79 (2004) 278–286.

[69] J. Kniewald, et al., Effect of s-triazine compounds on testosterone metabolism in the rat prostate, J. Appl. Toxicol. 15 (3) (1995) 215–218.

[70] J. Kniewald, P. Mildner, Z. Kniewald, Effects of s-triazine herbicides on hormone-receptor complex formation, 5 reductase and 3 hydroxysteroid dehydrogenase activity at the anterior pitiuitary level, J. Steroid Biochem. 11 (1) (1979) 833–838.

[71] B.J. Danzo, Environmental xenobiotics may disrupt normal endocrine function by interfering with the binding of physiological ligands to steroid receptors and binding proteins, Environ. Health Perspect. 105 (3) (1997) 294–301.

[72] J. Kniewald, P. Mildner, Z. Kniewald, Effects of s-triazine herbicides on 5-alpha dihydrotestosterone receptor complex formation in hypothalamus and ventral prostate, in: E. Action, F. Genazzani, W.I.P. DiCarlo, Mainwaring (Eds.), Pharmacological Modulation of Steroid Action, Raven Press, New York, 1980, pp. 159–169.

[73] F. Orton, et al., Endocrine disrupting effects of herbicides and pentachlorophenol: In vitro and in vivo evidence, Environmental Science & Technology 43 (6) (2009) 2144–2150.

[74] K. Pogrmic, et al., Atrazine oral exposure of peripubertal male rats downregulates steroidogenesis gene expression in Leydig cells, Toxicol. Sci. 111 (1) (2009) 189–197.

[75] K. Pogrmic, et al., Assesment of in vitro effects of atrazine on peripubertal rat Leydig cell steroidogenesis, FEBS J. 275 (2008) 462–1462.

[76] M.J.P. Fraites, et al., Characterization of the hypothalamic–pituitary–adrenal axis response to atrazine and metabolites in the female rat, Toxicol. Sci. 112 (1) (2009) 88–99.

[77] S.C. Laws, et al., Chlorotriazine herbicides and metabolites activate an ACTHdependent release of corticosterone in male Wistar rats, Toxicol. Sci. 112 (1) (2009) 78–87.

[78] L.M. Zorrilla, E.K. Gibson, T.E. Stoker, The effects of simazine, a chlorotriazine herbicide, on pubertal development in the female Wistar rat, Reproductive Toxicology 29 (4) (2010) 393–400.

[79] T.E. Stoker, C.L. Robinette, R.L. Cooper, Maternal exposure to atrazine during lactation suppresses suckling-induced prolactin release and results in prostatitis in the adult offspring, Toxicol. Sci. 52 (1999) 68–79.

[80] M. Roberge, H. Hakk, G. Larsen, Atrazine is a competitive inhibitor of phosphodiesterase but does not affect the estrogen receptor, Toxicol. Letters 154 (2004) 61–68.

[81] W. Fan, et al., Herbicide atrazine activates SF-1 by direct affinity and concomitant co-activators recruitments to induce aromatase expression via promoter II, Biochem. Biophys. Res. Commun. 355 (4) (2007) 1012–1018.

[82] W. Fan, et al., Atrazine-induced aromatase expression is SF-1-dependent: implications for endocrine disruption in wildlife and reproductive cancers in humans, Environ. Health Perspect. 115 (5) (2007) 720–727.

[83] A.C. Holloway, et al., Atrazine-induced changes in aromatase activity in estrogen sensitive target tissues, J. Appl. Toxicol. 28 (3) (2008) 260–270.

[84] F. Piferrer, Endocrine sex control strategies for the feminization of teleost fish, Aquaculture 197 (1–4) (2001) 229–281.

[85] Y. Guiguen, et al., Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish, Gen. Comp. Endocrinol. 165 (3 (Sp. Iss. SI)) (2010) 352–366.

[86] T.B. Hayes, Hormonal regulation of sex differentiation in amphibians, reptiles, and birds, in: E. Knobil, J. Neill (Eds.), Encyclopedia of Reproduction, Academic Press, San Diego, 1998, pp. 102–109.

[87] T.B. Hayes, Sex determination and primary sex differentiation in amphibians, J. Exp. Zool. 281 (1998) 373–399.

[88] D. Crews, Temperature-dependent sex determination: the interplay of steroid hormones and temperature, Zool. Sci. 13 (1) (1996) 1–13.

[89] W.H.N. Gutzke, J.J. Bull, Steroid hormones reverse sex in turtles, Gen. Comp. Endocrinol. 64 (3) (1986) 368–372.

[90] T.B. Hayes, Comparative endocrinology: a tool for understanding mechanisms and predicting effects of endocrine mimicking xenobiotics, in: L.J. Guillette (Ed.), Environmental Endocrine Disruptors: An Evolutionary Perspective, Taylor and Francis, NY, 1999.

[91] K.W. Wilhelms, et al., In ovo exposure to a triazine herbicide: Effects of atrazine on circulating reproductive hormones and gonadal histology in young Japanese quail, Archives of Environmental Contamination and Toxicology 51 (1) (2006) 117–122.

[92] S. Matsushita, et al., Effects of in ovo exposure to imazalil and atrazine on sexual differentiation in chick gonads, Poultry Science 85 (9) (2006) 1641–1647.

[93] S. Swan, et al., Semen quality in relation to biomarkers of pesticide exposure, Environ. Health Perspect. 111 (12) (2003) 1478–1484.

[94] J. Rayner, R. Enoch, S. Fenton, Adverse effects of prenatal exposure to atrazine during a critical period of mammary gland growth, Toxicol. Sci. 87 (1) (2005) 255–266.

[95] J. Rayner, C. Wood, S. Fenton, Exposure parameters necessary for delayed puberty and mammary gland development in Long–Evans rats exposed in utero to atrazine, Toxicol. Appl. Pharmacol. 195 (23–34) (2004).

[96] J.L. Rayner, et al., Atrazine-induced reproductive tract alterations after transplacental and/or lactational exposure in male Long–Evans rats, Toxicol. Appl. Pharmacol. 218 (3) (2007) 238–248.

[97] A. Pintér, et al., Long-term carcinogenecity bioassay of the herbicide atrazine in F344 rats, Neoplasma 37 (1980) 533–544.

[98] M. Ueda, et al., Possible enhancing effects of atrazine on growth of 7,12dimethylbenz(a) anthracene induced mammary tumors in ovariectomized Sprague–Dawley rats, Cancer Sci. 96 (1) (2005) 19–25.

[99] J. Eldridge, et al., Factors affecting mammary tumor incidence in chlorotriazine-treated female rats: hormonal properties, dosage, and animal strain, Environ. Health Perspect. 102 (11) (1994) 29–36.

[100] J. Stevens, et al., Hypothesis for mammary tumorigenesis in Sprague–Dawley rats exposed to certain triazine herbicides, J. Toxicol. Environ. Health. 43 (1994) 139–154.

[101] L.T. Wetzel, et al., Chronic effects of atrazine on estrus and mammary gland formation in female Sprague–Dawley and Fischer-344 rats, J. Toxicol. Environ. Health. 43 (1994) 169–182.

[102] M.A. Kettles, et al., Triazine exposure and breast cancer incidence: an ecologic study of Kentucky counties, Environ. Health Perspect. 105 (11) (1997) 1222–1227.

[103] M. Brodkin, et al., Atrazine is an immune disruptor in adult northern leopard frogs (Rana pipiens), Environ. Toxicol. Chem. 26 (1) (2007) 80–84.

[104] C. Cantemir, et al., p53 protein expression in peripheral lymphocytes from atrazine chronically intoxicated rats, Toxicol. Lett. 93 (2–3) (1987) 87–94.

[105] M.S. Christin, et al., Effects of agricultural pesticides on the immune system of Rana pipiens and on its resistance to parasitic infection, Environ. Toxicol. Chem. 22 (5) (2003) 1127–1133.

[106] M.S. Christin, et al., Effects of agricultural pesticides on the immune system of Xenopus laevis and Rana pipiens, Aquat. Toxicol. 67 (1) (2004) 33–43.

[107] N. Filipov, et al., Immunotoxic effects of short-term atrazine exposure in young male C57BL/6 mice, Toxicol. Sci. 86 (2) (2005).

[108] D. Forson, A. Storfer, Atrazine increases Ranavirus susceptibility in the tiger salamander Ambystoma tigrinum, Ecol. Appl. 16 (6) (2006) 2325–2332.

[109] D. Forson, A. Storfer, Effects of atrazine and iridovirus infection on survival and life history traits of the long-toed salamander (*Ambystoma macrodatylum*), Environ. Toxicol. Chem. 25 (1) (2006) 168–173.

[110] A.D. Gendron, et al., Exposure of leopard frogs to a pesticide mixture affects life history characteristics of the lungworm Rhabdias ranae, Oecologia 135 (3) (2003) 469–476.

[111] R. Hooghe, S. Devos, E. Hooghe-Peters, Effects of selected herbicides on cytokine production in vitro, Life Sci. 66 (26) (2000) 2519–2525.

[112] A. Houck, S. Sessions, Could atrazine affect the immune system of the frog Rana pipiens? Bios 77 (4) (2006) 107–112.

[113] N.A. Karrow, et al., Oral exposure to atrazine modulates cell-mediated immune function and decreases host resistance to the B16F10 tumor model in female B6C3F1 mice, Toxicology 209 (1) (2005) 15–28.

[114] K.R. Kim, et al., Immune alterations in mice exposed to the herbicide simazine, J. Toxicol. Environ. Health A: Curr. Issues 66 (12) (2003) 1159–1173.

[115] A.J. Langerveld, et al., Chronic exposure to high levels of atrazine alters expression of genes that regulate immune and growth-related functions in developing Xenopus laevis tadpoles, Environ. Res. 109 (4) (2009) 379–389.

[116] A. Liskova, et al., Effect of the herbicide atrazine on some immune parameters in mice, J. Trace Micropr. Tech. 18 (2) (2000) 235–240.

[117] K. Mizota, H. Ueda, Endocrine disrupting chemical atrazine causes degranulation through G(q/11) protein-coupled neurosteroid receptor in mast cells, Toxicol. Sci. 90 (2) (2006) 362–368.

[118] J. Pistl, et al., Determination of the immunotoxic potential of pesticides on functional activity of sheep leukocytes in vitro, Toxicology 188 (1) (2003) 73–81.

[119] W. Porter, J. Jaeger, I. Carlson, Endocrine, immune, and behavioral effects of aldicarb (carbamate), atrazine (triazine) and nitrate fertilizer mixtures at groundwater concentrations, Toxicol. Ind. Health 15 (1999) 133–150.

[120] S. Pruett, et al., Modeling and predicting immunological effects of chemical stressors: characterization of a quantitative biomarker for immunological changes caused by atrazine and ethanol, Toxicol. Sci. 75 (2) (2003) 343–354.

[121] A. Rooney, R. Matulka, R. Luebke, Developmental atrazine exposure suppresses immune function in male, but not female Sprague–Dawley rats, Toxicol. Sci. 76 (2003) 366–375.

[122] A. Rowe, et al., Immunomodulatory effects of maternal atrazine exposure on male Balb/c mice, Toxicol. Appl. Pharmacol. 214 (1) (2006) 69–77.

[123] A.M. Rowe, K.M. Brundage, J.B. Barnett, Developmental immunotoxicity of atrazine in rodents, Basic Clin. Pharmacol. Toxicol. 102 (2) (2008) 139–145.

[124] A. Rymuszka, et al., Application of immunostimulants after suppression induced byxenobiotics: effect of lysozyme dimer (KLP-602) after immunosuppression induced by atrazine in rabbits, Pol. J. Food Nutr. Sci. 13 (2) (2004) 51–54.

[125] M. Whalen, et al., Immunomodulation of human natural killer cell cytotoxic function by triazine and carbamate pesticides, Chem.-Biol. Interact. 145 (3) (2003) 311–319.

[126] D. Zeljezic, et al., Evaluation of DNA damage induced by atrazine and atrazinebased herbicide in human lymphocytes in vitro using a comet and DNA diffusion assay, Toxicol. In Vitro 20 (6) (2006) 923–935.

[127] J.R. Rohr, et al., Agrochemicals increase trematode infections in a declining amphibian species, Nature 455 (7217) (2008) 1235–1239.

[128] G. Giusi, et al., The endocrine disruptor atrazine accounts for a dimorphic somatostatinergic neuronal expression pattern in mice, Toxicol. Sci. 89 (1) (2006) 257–264.

[129] M.M. Hossain, N.M. Filipov, Alteration of dopamine uptake into rat striatal vesicles and synaptosomes caused by an in vitro exposure to atrazine and some of its metabolites, Toxicology 248 (1) (2008) 52–58.

[130] C.J. Martyniuk, et al., Aquatic contaminants alter genes involved in neurotransmitter synthesis and gonadotropin release in largemouth bass, Aquat. Toxicol. 95 (1) (2009) 1–9.

[131] V. Rodriguez, M. Thiruchelvam, D. Cory-Slechta, Sustained exposure to the widely used herbicide atrazine: altered function and loss of neurons in brain monoamine systems, Environ. Health Perspect 113 (6) (2005) 708–715.

[132] K.B. Tierney, et al., Relating olfactory neurotoxicity to altered olfactorymediated behaviors in rainbow trout exposed to three currently-used pesticides, Aquat. Toxicol. 81 (1) (2007) 55–64.

[133] M.D. Alvarez, L.A. Fuiman, Environmental levels of atrazine and its degradation products impair survival skills and growth of red drum larvae, Aquat. Toxicol. 74 (3) (2005) 229–241.

[134] F. de Noyelles, W. Kettle, D. Sinn, The response of plankton communities in experimental ponds to atrazine, the most heavily used pesticide in the United States, Ecology 63 (1982) 1285–1293.

[135] S.L. Dewey, Effects of the herbicide atrazine on aquatic insect community structure and emergence, Ecology 67 (1) (1986) 148–162.

[136] J.L. Griggs, L.K. Belden, Effects of atrazine and metolachlor on the survivorship and infectivity of Echinostoma trivolvis trematode cercariae, Arch. Environ. Contam. Toxicol. 54 (2) (2008) 195–202.

[137] J. Koprivnikar, R.L. Baker, M.R. Forbes, Environmental factors influencing community composition of gastropods and their trematode parasites in southern Ontario, J. Parasitol. 93 (2007) 992–998.

[138] J. Koprivnikar, M.R. Forbes, R.L. Baker, Contaminant effects on host-parasite interactions: atrazine, frogs, and trematodes, Environ. Toxicol. Chem. 26 (2007) 2166–2170.

[139] J. Rohr, P. Crumrine, Effects of an herbicide and an insecticide on pond community structure and processes, Ecol. Appl. 15 (2005) 1135–1147.

[140] J.R. Rohr, et al., Understanding the net effects of pesticides on amphibian trematode infections, Ecol. Appl. 18 (7) (2008) 1743–1753.

[141] J.R. Rohr, et al., Parasites, info-disruption, and the ecology of fear, Oecologia 159 (2) (2009) 447–454.

[142] J. Russo, L. Lagadic, Effects of parasitism and pesticide exposure on characteristics and functions of hemocyte populations in the freshwater snail *Lymnaea palustris* (*Gastropoda Pulmonata*), Cell Biol. Toxicol. 16 (1) (2000) 15–30.

doi: 10.5772/21282

Kinesiological Electromyography*

Vladimir Medved[1], Mario Cifrek[2]

[1]University of Zagreb, Faculty of Kinesiology
[2]University of Zagreb, Faculty of Electrical Engineering and Computing

Multiple rigid body modelling paradigm in the representation of human body and the inverse dynamic approach are shortly introduced first. Faithfull computer-based modelling of neuro-musculo-skeletal system is referred to next, as exemplified by one among the most successful approaches by Delp, introduced at the beginning of 1990es, and broadly implemented since.

The method of surface electromyography (sEMG) is explained, including concise description of the genesis of the signal, modelling the resulting "interference pattern" (the term by Basmajian and De Luca), and defining technical method of signal detection, amplification, conditioning, and interfacing to computer via analogue-to-digital (A/D) conversion.

Signal smoothing (or averaging), comprising full-wave rectification followed by low pass filtering, is the most important among time domain signal processing methods. This kind of signal representation bears resemblance to isometric muscular force signal – not available in principle – and can therefore be used as an indirect measure of muscular force. A number of kinesiological studies were realized in the past employing this rather non-invasive and elegant methodology to monitor muscular force(s). Applications range from medical rehabilitation examples, such as typically gait analysis, over studies of sportive movement patterns, to various ergonomic tasks. Processing multichannel sEMG signals may serve to quantify muscular coordination, enabling, in turn, evaluation of movement skill.

*The chapter was originally published in the book „Biomechanics in Applications/Book 1", edited by Václav Klika (V. Medved, and M. Cifrek, Kinesiological Electromyography), Rijeka, by the publisher InTech, 2011., 349-366, ISBN 978-953-307-969-1. This particular text was modified by an added Abstract. The article has been approved for scholarly and non-commercial use.

EMG signal spectrum is sensitive to local muscle fatigue, providing motivation for signal processing methods in the frequency domain. Numerous studies were performed yielding signal-based quantitative criteria of fatigue in various, primarily static, but also in dynamic tasks. We refer to classical and modern signal processing methods. Biomechanical applications, research or clinical alike, incorporating examples from the authors' own research where appropriate illustrate the methods.

We refer to state-of-the art engineering and technological solutions of multichannel telemetric sEMG recording systems, envisaging their possible novel future applications in kinesiology.

Key words:

Surface electromyography (sEMG), biomechanics, human movement, signal processing, kinesiology.

1. Introduction

Sometimes even identified – albeit incorrectly – with biomechanics, kinesiology is a field relying heavily on biomechanical methodology. Borellian Rennaisance approach, enhanced in the past with seminal contributions by scientists such as Marey, Braune, and Fischer, followed further by the work of the Berkeley Group, and later by a number of modern authors, has put classical mechanics in the centre of a paradigm taken to understand, analyse and quantitatively assess human movement. As this framework sets both a geometrical and a dynamical definition of the spatial (three dimensional – 3D) movement of human body as a whole, an important further focus of the study may be directed to skeletal muscle itself, a basic actuator of movement and genuine biological system designed to produce mechanical force and cause movement. In this context, to monitor and evaluate human movement, we have a unique, second to none, method: electromyography (EMG); i.e. the recording of electrical activity of skeletal musculature. When studying kinesiological tasks, in particular, surface electromyography (sEMG) is the method's variant of choice. To quote Hess (Hess, 1954, as cited in Waterland, 1968): „The course of a movement is nothing else but a projection to the outside of a pattern of excitation taking place at a corresponding setting in the central nervous system". This thought reflects the importance of EMG signals as certain "windows" into the action of the central nervous system during the performance of a motor task.

Kinesiological electromyography is, therefore, an established subfield of modern locomotion biomechanics. We witness today a number of professional journals, conferences, organizations and university-level courses devoted to this subject around the world. At the University of Zagreb, in particular, courses of this kind are

spread across several departments (Medved, 2007). At the intersection of physiology and biomechanics, and with strong quantitative aspect, this discipline significantly contributes to our understanding of human movement and is therefore used in a number of basic and applied fields. Consequently, due to its inter-disciplinary nature, it is used by different professionals; physical therapists, medical doctors of various specialties, electrical and biomedical engineers, kinesiologists, to name but a few.

To set the global framework, this chapter first shortly refers to the basics of methodology in biomechanical approach to human movement and of musculo-skeletal modelling. The method of surface electromyography is described next in some detail. Time domain signal processing methods are then presented, followed by the methods performed in the frequency domain; all with the vision towards applications in the field of kinesiology. Chapter concludes by pointing to modern engineering solutions for multichannel sEMG.

2. Multiple rigid body paradigm and neuro-muscular modelling

Biomechanical methodology for studying human movement is based on a multiple rigid body modelling paradigm in the representation of human body. Body segments are presumed to be rigid and interconnected by joints, so that a so-called kinematic chain is formed as a relevant description of movement of the body as a whole. At this level of abstraction (and simplification), the laws of classical mechanics may be applied to this system, whereby experimentally obtained (measured) kinematic data are combined with inertial properties of body segments, taking into account possible external acting forces and moments. The strict expression of quantitative relations in a system of this kind enables mathematical calculation of internal resultant (net) forces and moments acting in virtual joint centres, the procedure called inverse dynamic approach. Historically, the approach was first introduced by Braune and Fischer in Germany near to the end of 19th century, who have employed the method of photography to realize stereometry in movement recording and, consequently, implemented relevant Newtonian equations. The approach was later refined and perfected as technology for measurement and signal and data processing developed to our days. A detailed overview of the subject matter including exact mathematical formalisms describing the approach, as well as descriptions of a number of practical solutions, is available in representative journal papers (Cappozzo, 1984; a landmark paper) and in standard biomechanics literature (Medved, 2001; Rose & Gamble, 2006). This methodology forms the basis of modern biomechanics of human movement and locomotion, both in sportive and in medical applications.

Inverse dynamic approach is thus, to recapitulate, a vehicle to obtain quantitative estimates of internal resultant (net) forces and moments acting in the joints during movement of the body, idealized as a multi-segment mechanical system.

Skeletal muscle was researched the most during the 20th century, at the microstructural, biophysical, biochemical as well as at the control level. This was enabled by technological and methodological advancements such as the invention of electron microscopy, development of electrophysiological recording techniques, pursuing the concept of cybernetics, etc. Basic muscle contraction mechanism was elucidated and mathematically modelled: see, for instance, explanation of the Huxley's model of muscle contraction (McMahon, 1984). Up to the 1980es, therefore, theoretical basis was available for development of faithful quantitative models of the muscle-tendon complex (Zajac, 1989), based on which computer-supported quantitative and graphics-based solutions to model the neuro-musculo-skeletal system were realized (Delp et al., 1990). The approach has been broadly implemented until our days, not only in research, but entering the arena of clinical applications as well (Delp et al., 2000).

This methodological improvement has added to the classical inverse dynamic approach, enabling, in fact, further sub-division of calculated resultant (net) forces and moments in the joints into their components corresponding to particular muscles, resulting with a detailed and realistic biomechanical modelling and simulation possibilities of complex neuro-musculo-skeletal structures.

Skeletal muscle is a system characterized by mechanical, thermal and electrical energy outputs. Mechanical action of skeletal muscle as a whole is described well by the „tension-length" and „force-velocity" relations, its model including active, elastic and viscous components (Medved, 2001). EMG-supplied information is connected with fundamental muscle function. The active component, namely, is the one representing genuine feature of muscular tissue to mechanically contract, and this component is correlated with electrical events; being manifested ultimately as electromyographic signals (provided adequate detection and recording be secured). Contributions of elastic and viscous components of the model to the muscle force, on the contrary, are not „visible" in the EMG.

3. Surface electromyography

Electromyography means detection and recording the electrical activity of skeletal musculature. In kinesiology, predominantly surface recording technique is used due to the requirement of non-invasiveness. To correctly comprehend the method of

electromyography, a certain level of understanding of signal genesis is necessary. Based on anatomico-physiological properties of neural and muscular tissues, the process may be mathematically modelled; a task accomplished successfully by Carlo De Luca, electrical and biomedical engineer, who in 1960es and 1970es gave a careful and systematic mathematical description of a so-called interference pattern – a resulting global electrical signal by the active muscle as a whole – and thus complemented the traditional anatomically-based approach. Interested reader is referred to original papers (De Luca, 1979, 1984) as well as to the book „*Muscles Alive: Their Functions Revealed by Electromyography*" (Basmajian & De Luca, 1985) – a classical reference in the field – where mathematical modelling of interference pattern is also reproduced.

3.1 On origin and properties of myoelectrical signal

Processes of depolarization and repolarization result with action potentials at the muscle fibre membrane. The depolarization–repolarization cycle forms a depolarization wave or electrical dipole travelling along the surface of a muscle fibre. Since a motor unit consists of a number muscle fibres, the electrode pair (detection electrodes issues will be discussed later) "sees" the potentials of all active fibres within this motor unit, depending on their spatial distance from the detection site. Typically, the action potentials sum up to a so-called Motor Unit Action Potential (MUAP), which differs in form and size depending on the geometrical fibre orientation with respect to the electrode(s) site. Within kinesiological studies, the MUAPs of all active motor units detectable under the electrode(s) site are electrically superimposed and observed as a bipolar signal with symmetric distribution of positive and negative amplitudes (mean value equals to zero). This is interference pattern (Konrad, 2005).

An unfiltered (exception: amplifier bandpass) and unprocessed signal comprising the superimposed MUAPs is called a raw EMG signal. In Fig. 1, a raw surface EMG (sEMG) recording is shown for three successive contractions of m. *rectus femoris*.

Raw EMG signal is, by its nature, of random shape (quasi-stochastic), meaning that one raw recording burst cannot be precisely reproduced in exact shape. This is due to the fact that the actual set of recruited motor units constantly changes within the matrix of available motor units: If occasionally two or more motor units fire at the same time, and they are located near the electrodes, they produce a strong superposition spike. By applying a smoothing algorithm (e.g. moving average) or yielding a proper amplitude parameter (e.g. area under the rectified curve), the non-reproducible contents of the signal is minimized.

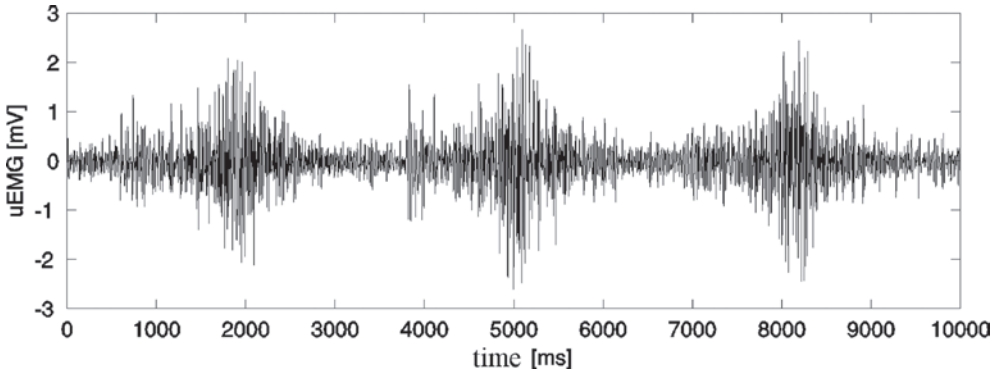


Fig. 1 – Raw surface EMG recording for three successive contractions of m. *rectus femoris* (Cifrek, 1997).

Raw sEMG can range between ±5 mV (maximum achieved in athletes) and typically the frequency contents ranges between 6 and 500 Hz, showing most power between ~20 and 250 Hz.

3.2 Measurement of surface EMG signal

In majority of kinesiological studies surface electrodes are used due to their non-invasiveness. Offering the benefit of easy handling, their main limitation is that only surface muscles can be detected. For deeper muscles (covered by surface musculature or bones) fine-wire or needle electrodes are inevitable. (Fine-wire electrodes, being thin and flexible, are better suited to kinesiological applications than needle electrodes.)

Surface EMG electrodes can be classified considering the materials and the technologies adopted for their manufacturing (Merletti et al., 2009). One can distinguish between dry and non-dry or wet electrodes. Several types of dry electrodes exist: pin or bar electrodes made of noble metals (e.g. gold, platinum or silver), carbon electrodes, and sintered silver or silver chloride electrodes. Wet electrodes include a layer of conductive gel, hydrogel or sponge saturated with an electrolyte solution. These electrodes are often self-adhesive, so they can be easily applied and used for analysis of dynamic sEMG (Merletti et al., 2009).

Among surface electrodes, silver/silver chloride pre-gelled electrodes are the most often used ones and recommended for the general use (SENIAM, according to Hermens et al., 1999). The electrode diameter (conductive area) should be sized to 1 cm or smaller. Commercial disposable electrodes are manufactured as wet gel electrodes or adhesive gel electrodes. Generally wet gel electrodes have better conduction and impedance conditions (i.e. lower impedance) than adhesive gel electrodes. The latter one has the advantage that it can be repositioned in case of errors.

Electrodes are positioned in a so-called differential arrangement; meaning that to each specific skeletal muscle pair of electrodes is to be attached according to the standardized procedure regarding their location with respect to the muscle, and with standard spacing. It is common today to follow the SENIAM standards (Hermens et al., 1999). There is an on-going debate among the experts, however, regarding the actual positioning of the electrodes with regard to muscle for kinesiological measurements. The conservative opinion regarding two signal electrodes was that they have to be positioned at the midpoint, the most prominent part of muscle, at a distance of 15 to 30 mm (Nilsson et al., 1985). A more exact approach to electrode positioning, however, presupposes that the location of the motor point (plate) has been determined beforehand. This is accomplished by electrically stimulating the muscle and determining the location of stimulation where the muscle has the greatest mechanical response. For a long time the opinion held was that electrodes should be positioned as close as possible to the motor point (Viitasalo et al., 1980). Loeb and Gans explain this traditional attitude; they think that "electrodes have to be positioned reasonably close to the motor point with the goal of obtaining a signal of maximum and constant amplitude" (Loeb & Gans, 1986). But, from the point of view of signal stability, this location is the worst. In this region the action potentials travel caudally and rostrally along muscular tissue consisting of fibres, and so the positive and negative phases of the action potential are mutually neutralized. Basmajian and De Luca, therefore, are of the opinion that electrodes must be located approximately at the midpoint between the determined motor point location and the point where the muscle and tendon join because there signal properties are the most stable.

As far as interelectrode distance is concerned, De Luca and Knaflitz recommend a value of 10 mm centre to centre (De Luca & Knaflitz, 1992). Namely, the interelectrode distance influences signal spectrum (Lynn et al, 1978). It is therefore necessary to keep the distance fixed, so as to enable quantitative comparisons of measured values intra and inter-muscularly, as well as between subjects. A 10 mm distance is considered to be a good technical compromise because in this way a representative electrical muscle activity is detected during contraction (several cm^3 of muscular tissue), while the filtering effect of bipolar configuration is reduced at the same time (Lindström, 1973). But, in measuring dynamical muscle activity, it is often impossible to keep the interelectrode distance constant, introducing additional variability to the measurement procedure. It is customary to locate the third, neutral electrode as far away as possible from the muscle. As an addendum to experiment documentation, and in order to achieve repeatability of the measurement procedure, it is a common practice to take a photograph of the actual electrode setting. The dilemmas mentioned remain open. Besides the mentioned SENIAM protocol, valuable are also the Standards for Reporting EMG data (*Journal of Electromyography and Kinesiology*, February 1999; 9(1):III-IV).

After detection follow signal amplification and conditioning, bringing the signal into the volt range. Pre-amplifier is positioned as close as possible to detection site and is of a differential type. An important property of differential amplifier is high quality signal amplification with simultaneous suppression of noise (Medved, 2001). Interfaced to computer via analogue-to-digital (A/D) conversion, signals may be digitally processed, the task which can be accomplished either in the time domain or in the frequency domain.

Modern electromyograph devices secure high-quality signal recording with good noise suppression. Usually, they are designed as data-loggers or, alternatively, radiotelemetric systems (for example ZeroWire by Noraxon, FREEEMG 300 by BTS). Considering problems of muscle coordination and synchronisation when performing movement patterns, it is desirable to measure more EMG channels simultaneously. Modern electronics technology enables small and light detection-amplification devices as well as reliable signal transmission. Displaying a multichannel sEMG signal series provides a visually attractive means to monitor muscular activity, serving, as a first step, in qualitative analysis of multiple muscle action, be it isometric or dynamic.

4. SEMG signal processing in time domain: A muscle co-ordination issue

There are several typical time domain signal processing methods used in electromyography. All of them aim to simplify quantification and, subsequent, interpretation of signals recorded (Medved, 2001). (Raw EMG signals are, namely, of a quasi-stochastic and noise-like appearance: Fig. 1.) Among the proposed quantification methods pretending to offer indices correlating well with muscle force and energy, the most important one from the kinesiological point of view is signal smoothing (or averaging) which comprises full-wave rectification followed by low pass filtering. This kind of signal representation bears resemblance to isometric muscular force signal – which, in principle, is not available – and can therefore in the first approximation be used as an indirect measure of muscular force (De Luca, 1997).

A number of kinesiological studies were realized in the past employing this rather non-invasive and elegant methodology to monitor muscular force(s). Spectrum of applications ranges from a number of medical rehabilitation examples, such as typically gait analysis (Fig. 2; see typical multichannel lower extremity EMG record of a walking child, Frigo & Crenna, 2009), over studies of sportive movement patterns, to various ergonomic problems. (Needless to say that a typical kinesiolog-

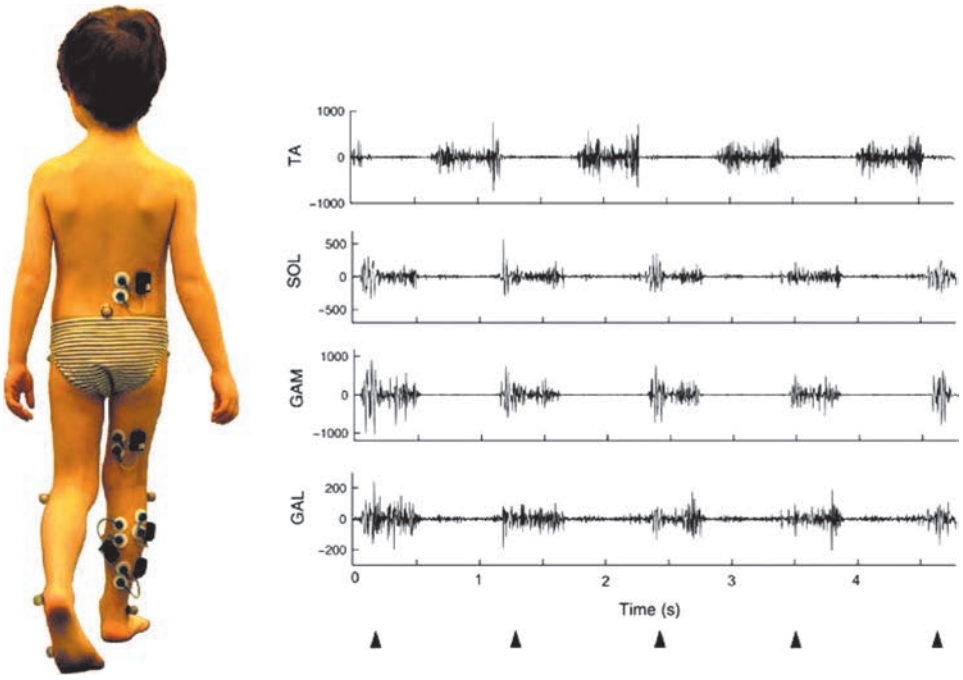


Fig. 2 – Wireless sEMG recording in a 5 years old child. The picture on the left shows the electrodes and the self-powered cases, each one provided with preamplifier and antenna for independent transmission of myoelectric signals. Traces on the right-side are illustrative examples of EMG activities recorded during a tiptoe walking task from *tibialis anterior* (TA), *soleus* (SOL), *gastrocnemius medialis* (GAM), and *gastrocnemius lateralis* (GAL) (Frigo & Crenna, 2009).

ical experimental study incorporates, besides EMG, also other measurement quantities: kinematic and kinetic, depending upon availability.)

Multichannel EMG may serve in studies of muscular coordination, enabling, in turn, certain evaluation of locomotor skill. An example of this kind of study whereby skilled artistic gymnastics movements were measured and analysed will be referred to (Medved & Tonković, 1991; Medved et al., 1995; reproduced in Medved, 2001). Gymnasts were instrumented with surface electrodes positioned at major lower extremity muscles (m. *gastrocnemius*, m. *tibialis anterior*, m. *rectus femoris*, m. *biceps femoris*). They were instructed to perform backward somersaults from the standing position with take-off from force platform (Fig. 3). Performances were graded by certified gymnastics judges. This gymnastics element of technique was chosen as it enabled insight into the level of acquired performance skill, because it concerns a complex movement structure. Gymnasts take a number of years of training to acquire a high-quality backward somersault and this element represents a significant component of the performance repertoire of their compositions.

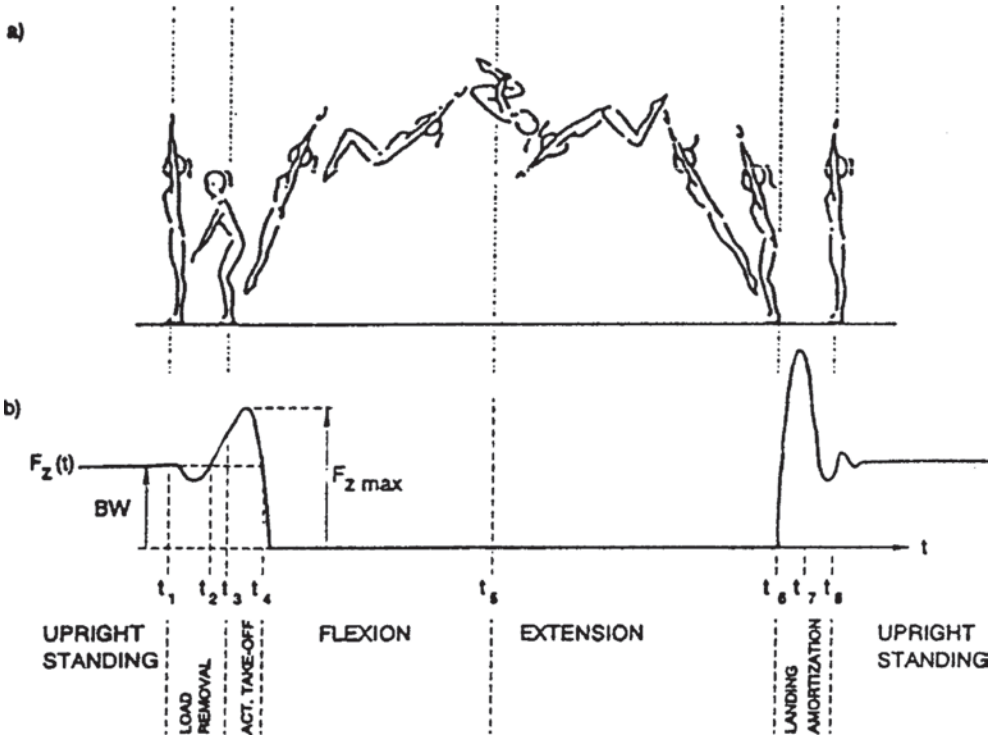


Fig. 3 – a) Schematic sequential representation of the backward somersault kinematics b) Idealized waveform of the vertical component of ground reaction force vector F_z (t), (BW= body weight, $F_{z\,max}$ = maximum value of vertical force signal). Force signal waveform may be correlated to movement kinematics during the time period preceding airborne phase (Medved et al., 1995).

SEMG signals were detected, amplified and averaged on-line by analogue means, i.e. full-wave rectified and low pass filtered (analogue RC filter, 100 ms time constant), which was a part of the RM Beckman Dynograph device ("averaged" measurement mode). The upper frequency thus attained was 150 Hz. Signals were further digitized on-line and stored in computer memory. Quantification according to Gandy et al., (1980) was applied. The above procedure was thus: 1) measurement and signal acquisition, 2) signal pre-processing (i.e. signal smoothing), 3) statistical signal and data processing (i.e. calculation of correlation between smoothed signals, and of correlation between signal parameters and grades of performance) and 4) conclusion, i.e. determining quantitative skill criteria.

The experiments yielded the following quantitative criteria for the level of skill acquisition in the performance of the backward somersault from a standing position. The kinetic criterion of good quality performance is determined by values of the vertical force Fz impulse width < 300 ms and of ratio Fzmax/BW > 3 (Fig. 3), while the bioelectric criterion is determined by the value of the correlation coeffi-

cient of averaged EMG signals of the left and right m. *gastrocnemius* of ≥ 0.8, reflecting a high degree of symmetry in the activity of ankle extensor muscles (BW stands for body weight).

The bioelectrical criterion has been further elaborated into the so-called moving correlation function:

$$H(j)=\frac{\sum_{i=1}^{200}\left[A(i+j)-\bar{A}\right]\cdot\left[B(i+j)-\bar{B}\right]}{\sqrt{\sum_{i=1}^{200}\left[A(i+j)-\bar{A}\right]^2\cdot\sum_{i=1}^{200}\left[B(i+j)-\bar{B}\right]^2}} \tag{1}$$

The function $H(j)$, being a collection of scaled correlation coefficients, calculated one by one for each j shows the correlation between two selected averaged EMG signals. It is calculated by moving a 200 point window A over the original 300 point function B starting from the " 50 point" to the "+50 point" (Schwartz, 1975; Spiegel, 1992). The function H(j) thus has 100 points in total (j = 100) with an expected maximum around or at the "50 point". Fig. 4 shows calculated moving

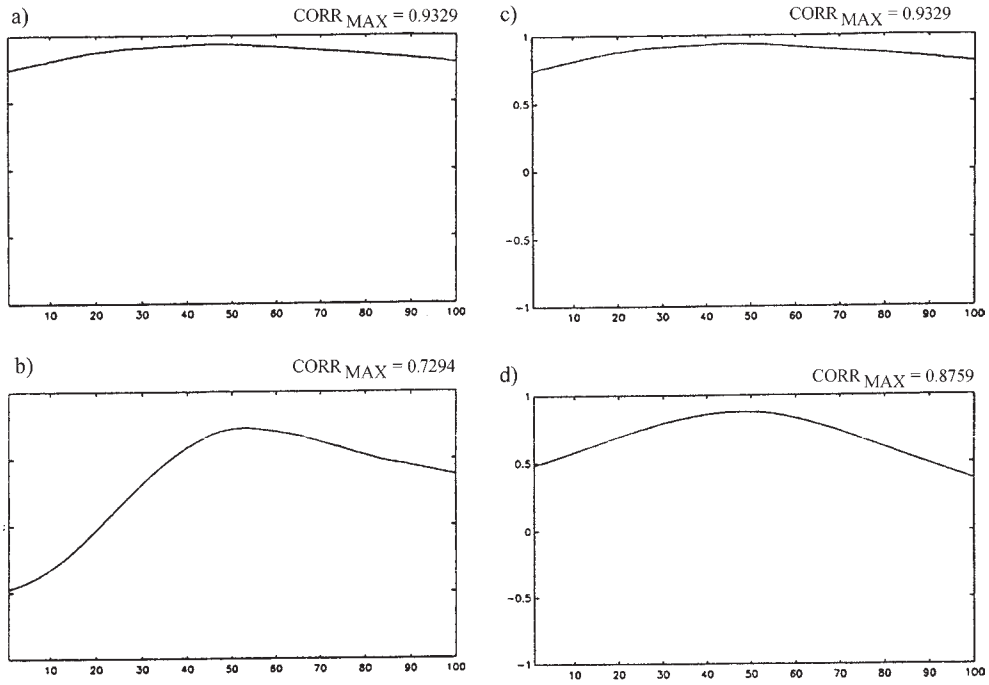


Fig. 4. – Left column: correlation (L Ga vs. R Ga) functions H(j) of "top-level" performer (a) and "poor" performer (b). Right column: correlation (L Ga vs. R Ga) functions H(j) of top-level gymnast performing backward somersault "at his best" (c) and "deliberately poorly" (d) (Medved et al., 1995).

correlation functions for the "top-level" and "poor" performer, that is, for performances by a top-level performer "at his best" and "deliberately poor", respectively. A good discriminability feature is observed in the procedure for the evaluation of skill level realized in this way; EMG signals have thus shown to be rather sensitive measures of neuro-muscular performance.

The method described serves as an example of possible use of multichannel sEMG signals as an indirect measure of multiple muscle force co-ordination pattern associated with particular skilled locomotion. It is potentially applicable in quantification of acquisition of other movement structures as well (presuming respective muscles are measured), and might also serve in monitoring the progress in motorics in course of particular diagnostics and/or treatment procedures in rehabilitation medicine.

Depending upon a kind of question attempted to be answered by EMG analysis, signal amplitude normalization might be necessary. This is for instance when inter-subject or inter-muscular (at the same subject) comparisons of EMG signals are to be made. Naturally, the value to which normalization is made (100%) must be determined precisely in a sense of defining actual kinesiological conditions of a corresponding movement or posture, and type of contraction performed (isometric, dynamic,...). There is no absolute consensus about this matter, however, and each investigator is responsible for correctness of his own measurement/experiment (Konrad, 2005; Medved, 2001).

5. SEMG signal processing in frequency domain: A muscle fatigue issue

Given the recommended amplifier bandpass settings from 10 Hz high-pass up to at least 500 Hz low pass (SENIAM), most of the surface EMG frequency power is located between 20 and 250 Hz. Power distribution can be obtained by the Fourier Transformation (applying in practice Fast Fourier Transform (FFT) to a time represented signal) and graphically presented as EMG signal power density spectrum, which shows signal power distribution with regard to frequency (Fig. 5).

The dominant change in the EMG power density spectrum during sustained contractions is a compression of the signal spectrum toward lower frequencies, which is shown by curves on Fig. 5 a) and b). Measures of this compression are associated with metabolic fatigue in the underlying muscle.

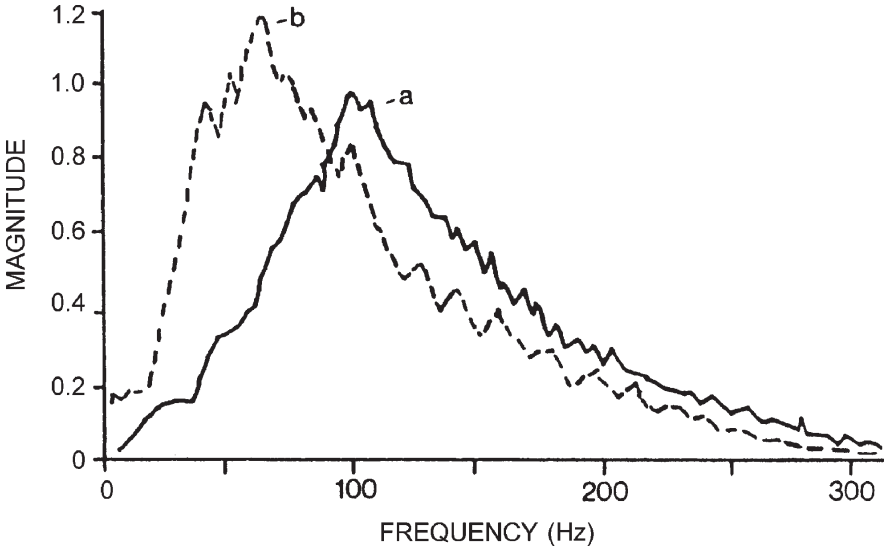


Fig. 5 – SEMG power spectrum density – before (a) and after (b) fatiguing exercise (De Luca, 1984).

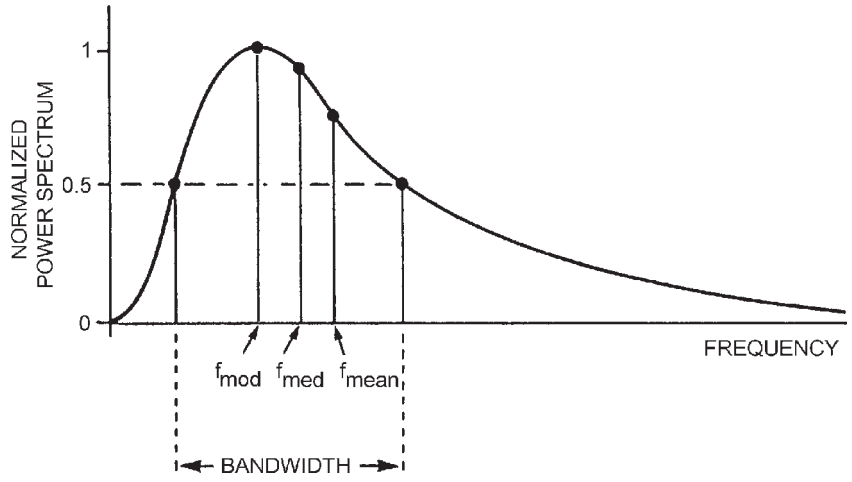


Fig. 6 – Power spectrum density characteristic frequencies

Power spectrum density curve can be characterized by the following frequency parameters (Fig. 6): mean frequency, as the mathematical mean of the spectrum curve:

$$f_{mean} = \frac{\int_0^{f_s/2} fP(f)\,df}{\int_0^{f_s/2} P(f)\,df} \tag{2}$$

and median frequency as the parameter that divides the total power area into two equal parts:

$$\int_0^{f_{med}} P(f)\,df = \frac{1}{2}\int_0^{f_s/2} P(f)\,df \tag{3}$$

Within applied EMG-frequency analysis the mean and median frequencies are the most important parameters, and their time domain changes in sustained contractions are monitored (fatigue studies). (An alternative to the FFT based calculations was, historically, the simple counting of zero line crossings of the EMG signal, being highly correlated to the FFT based mean/median frequency values.)

Within static submaximal contractions both amplitude and frequency based analysis parameters show time domain changes due to muscular fatigue. The classical test requires a constant load level at a well-defined angle position/muscular length. Due to recruitment of motor units, the amplitude shows an increase, whereas mean and median frequency of the power spectrum show a decrease over contraction time. The latter ones decline because – besides other reasons – the conduction velocity of the motor actions potentials at the muscle membrane decreases. This causes a shift to the left of the power density spectrum. The regression coefficient of the median or mean frequency slope towards lower values can be used as a non-invasive fatigue index for the investigated muscle.

The influence of muscle fatigue on the properties of the sEMG signal during isometric voluntary and electrically elicited contractions is clearly shown in Fig. 7 (Merletti & Lo Conte, 1997). In this example a subject maintained target torque level for 60 s before a mechanical manifestation of muscle fatigue occurred (healthy *tibialis anterior* muscle). Increase of the RMS value and decrease of CV and power spectrum mean frequency are evident from the beginning of the contraction. This is even more evident during electrically elicited contractions (vastus medialis stimulated for 30 s at 30 pulses/s), and it appears to be a combination of scaling (stretching in time and in amplitude) and a change of shape of the M wave (myoelectric signal evoked by electrical stimulation).

Cifrek and colleagues (Cifrek, 1997; Cifrek et al., 1998, 2000) developed a method of surface myoelectric signal measurement and analysis aimed at evaluating muscle fatigue in healthy subjects during cyclic dynamic contractions of upper leg musculature in a simple cyclic flexion–extension movement of the lower leg, recorded during exercise on a training device (Fig. 8). The signal processing part of the method is schematically presented in Fig. 9. As an indicator of muscle fatigue a change in the power spectrum median frequency (MF), calculated from the spectro-

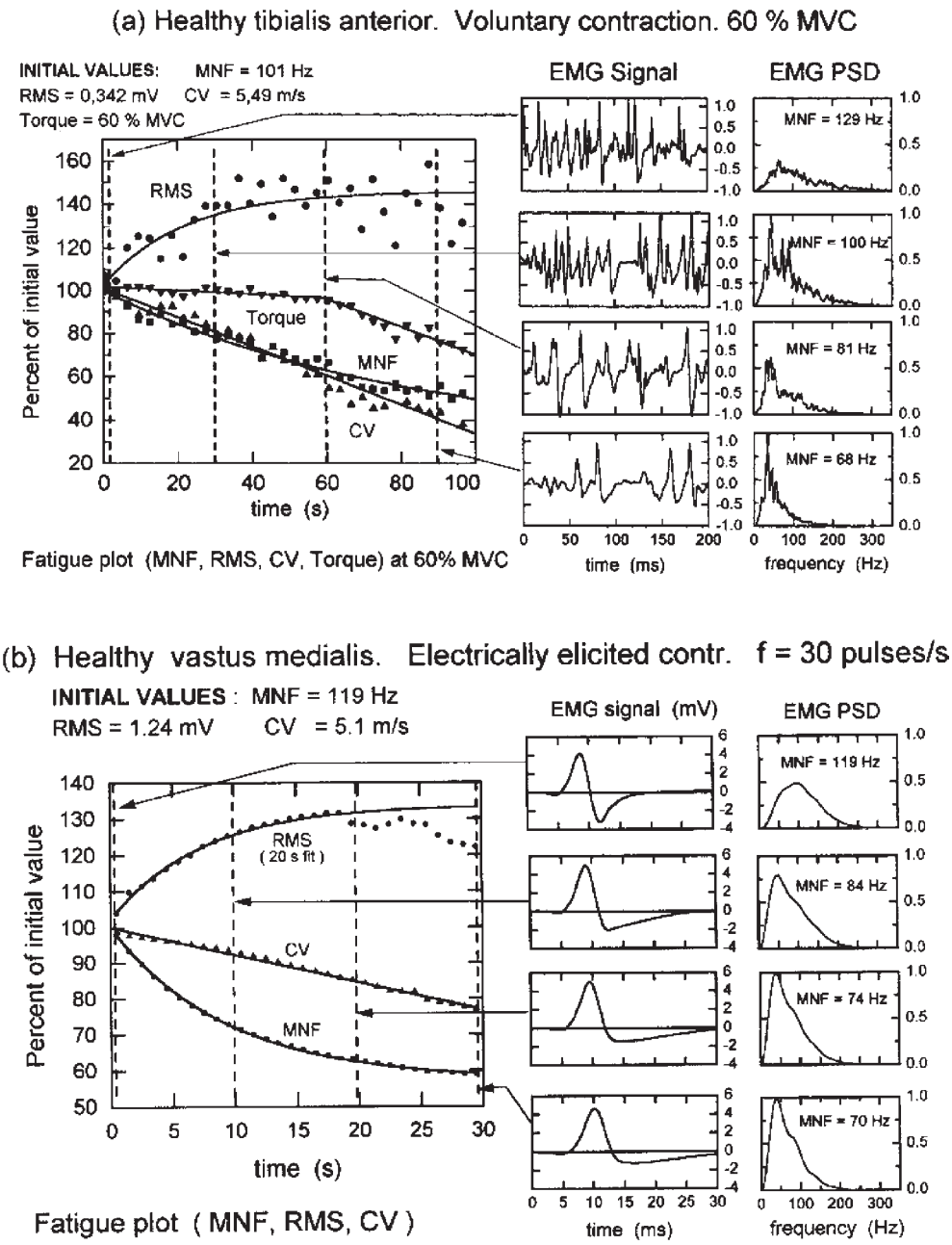


Fig. 7 – Examples of fatigue plots showing the time course of the EMG signal variables during a sustained contraction, the EMG signal and its power spectral density (PSD) during specific time windows. (a) Voluntary contraction of a healthy tibialis anterior muscle sustained for 100 s with a target set at 60% MVC. For the sake of clarity a three-point moving average has been applied to the variables and one value every 3 s is displayed. Note the mechanical breakpoint at 60 s. (b) Electrically elicited contraction of a healthy vastus medialis stimulated for 30 s at 30 pulses/s. MNF = mean frequency of the PSD, RMS = root mean square value, CV = conduction velocity. From (Merletti & Lo Conte, 1997).

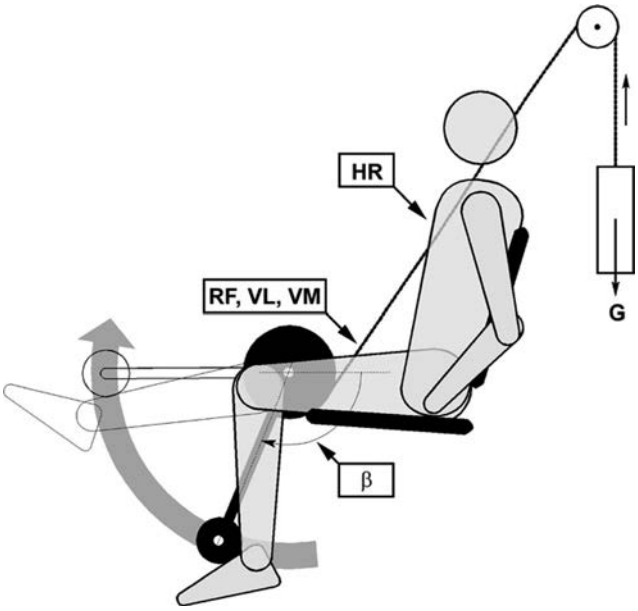


Fig. 8 – Exercise on a "leg-extension" training device and measured quantities (HR=heart rate, RF = m. *rectus femoris*, VL = m. *vastus lateralis*, VM = m. *vastus medialis*, β = shaft angle) (Cifrek et al., 2000).

gram, was used. The authors also discussed the influence of analysis parameters on the results (Cifrek et al., 1999).

Merletti & Parker (2004) have edited a book providing a broad coverage of modern modelling and signal processing issues in the area of sEMG, among other also fatigue influences and means of quantification of this phenomenon. Cifrek et al. (2009), however, have recently presented a state of the art summary on the issue of sEMG based muscle fatigue evaluation. An overview is given of classical and modern signal processing methods and techniques from the standpoint of applicability to sEMG signals in fatigue-inducing situations relevant to the broad field of biomechanics. Time domain, frequency domain, time-frequency and time-scale representations, and other methods such as fractal analysis and recurrence quantification analysis are described succinctly and are illustrated with their biomechanical applications, research or clinical alike.

SEMG recordings during dynamic contractions are particularly characterised by non-stationary (and non-linear) features. Standard signal processing methods using Fourier and wavelet based procedures demonstrate well known restrictions on time–frequency resolution and the ability to process non-stationary and/or non-linear time series, thus aggravating the spectral parameters estimation. The Hilbert–

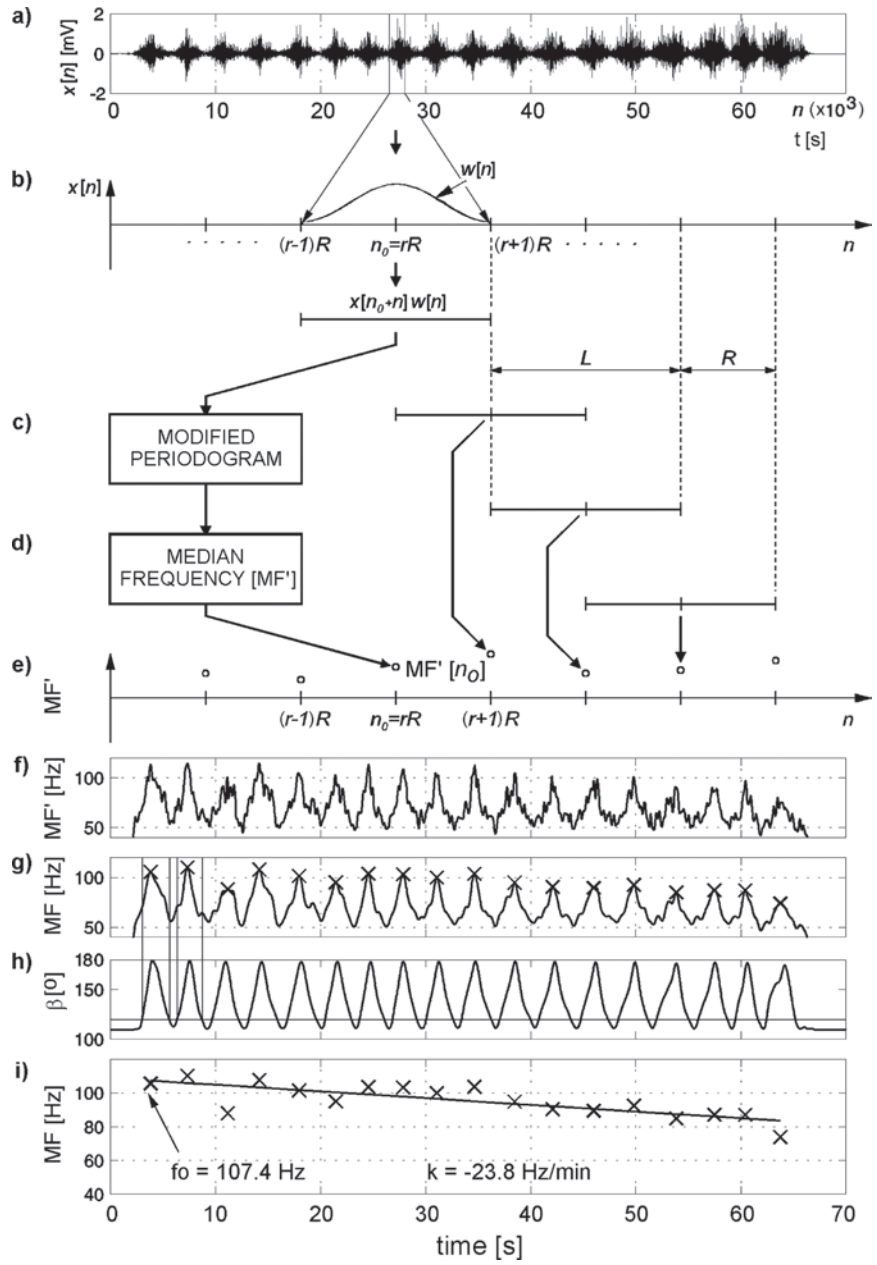


Fig. 9 – Myoelectric signal spectral analysis for quantification of muscle fatigue during dynamic contractions: (a) sEMG signal x[n], raw data; (b) extracted data, using window sequence w[n] of length L, with shift of R samples (c), (d) and (e) estimation of median frequency (MF') using modified periodogram of windowed sequence, (f) course of median frequency (MF'), (g) after low-pass filtering, maximum values of MF during each contraction were calculated, (h) limits of contractions have been calculated using shaft angle data, (i) a slope of the regression line (k, expressed in Hz/min) that fits maximum values of MF in a least-square sense was used as a fatigue index. From regression line, the frequency at the beginning of exercise (f_0) was calculated (Cifrek et al., 2000).

Huang transform (HHT), comprising of the empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA), provides a new approach to overcome these issues (Srhoj-Egekher at al., 2010). The time-dependent median frequency estimate is used as muscle fatigue indicator, and linear regression parameters are derived as fatigue quantifiers.

Moreover, emerging methods based on nonlinear signal analysis are being applied. These techniques, known as recurrence quantification analysis (RQA), are based on detecting deterministic structures in the signals that repeat throughout a contraction (Farina et al., 2002).

6. Conclusion

The presented methods of sEMG signal measurement and processing were based on a classical differential (bipolar) signal detection and amplification. Currently, improved measurement techniques, including multi-channel approaches targeted at a single muscle are being developed, shifting a focus from a one-dimensional signal-based considerations to two-dimensional surface-based approaches registering myoelectric phenomena (Fig. 10). Merging these new measurement possibilities

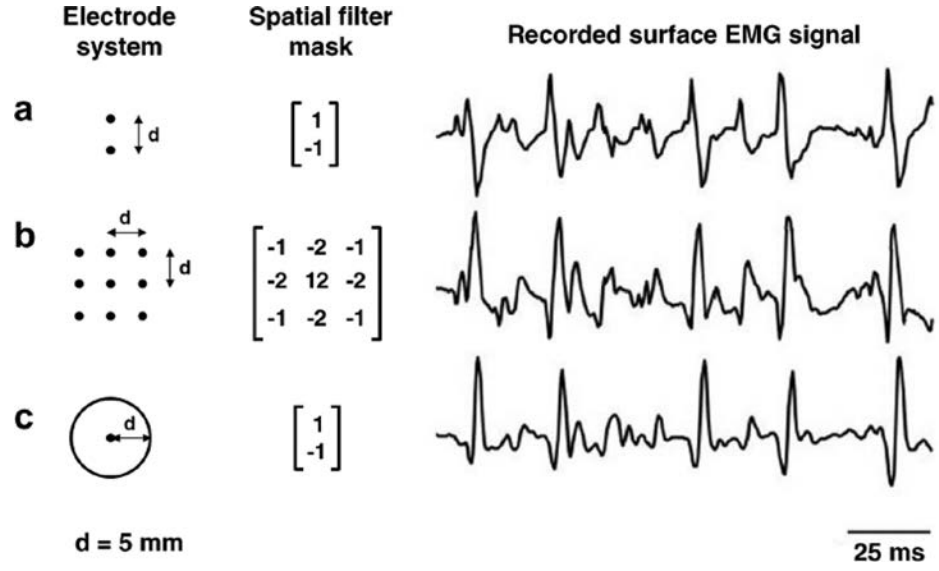


Fig. 10 – Examples of sEMG signal recorded, during the same contraction, with different spatial filter configurations: (a) single differential system; (b) inverse binomial filter of the second order (IB2); (c) single ring concentric electrode system. In all cases, the interelectrode distance was 5 mm. The greater spatial selectivity of the concentric electrode system with respect to the other systems is evident (Merletti et al., 2009, redrawn and adapted from Farina & Cescon, 2001).

with sophisticated mathematical methods and digital signal processing techniques provides a solid basis for validation, refinement and standardization of suitable new methods to be applied in biomechanical situations.

Methods for analyzing fatigue at the single motor unit level relying on non-invasive multichannel recordings and joint use of spatial filtering and spatial sampling are currently under study (Merletti et al., 2003; Farina et al., 2004).

On the other hand, we feel confident that in realms of biomechanical research, the presented methods for muscle fatigue evaluation will be further developed, exercised, improved and standardized. In clinical diagnostic applications – both in sport and in medical rehabilitation contexts – standardization of modern methods embodied in a novel type of a "muscle fatigue monitor" device is yet to be realized. It may appear in a form of a compact device of portable design and makeup, offering a menu of several correlated fatigue indices, (including, possibly, some non-EMG based as well). This goes in line with the general feature of miniaturisation of biomedical electronics instrumentation, enabling its use in an increasing number of real-life situations.

Acknowledgements

The results presented are the product of a number of scientific projects including *"Noninvasive measurements and procedures in biomedicine"*, *"Automated motion capture and expert evaluation in the study of locomotion"* and *"Real-life data measurement and characterization*", supported by The Ministry of Science, Education and Sports, Republic of Croatia.

References

[1] Basmajian, J.V. & De Luca, C.J. (1985). Muscles Alive: Their Functions Revealed by Electromyography. Fifth Edition, Williams & Wilkins, ISBN 0-683-00414-X, Baltimore, Md., USA

[2] Cappozzo, A. (1984). Gait Analysis Methodology. Human Movement Science, Vol. 3, pp. 27-50, ISSN 0167-9457

[3] Cifrek, M. (1997). Myoelectric Signal Analysis during Dynamic Fatigue. Ph.D. Dissertation. University of Zagreb, Faculty of Electrical Engineering and Computing (in Croatian).

[4] Cifrek, M., Tonković, S. & Medved, V. (1998). Surface Myoelectric Signal Spectral Analysis during Fatigued Dynamic Contractions of Quadriceps Muscle. In: R. Magjarević (Ed.), Biomedical Measurement and Instrumentation: Proceedings of the 8th International IMEKO Conference on Measurement in Clinical Medicine, Vol. 3, KoREMA, Zagreb, pp. 98–101, ISBN 953-6037-26-2

[5] Cifrek, M., Tonkovic, S. & Medved, V. (1999). Surface EMG Spectrogram in Dynamic Muscle Fatigue Monitoring – Influence of Analysis Parameters. in: H. Hinrikus et al. (Eds.), 11th Nordic-Baltic Conference on Biomedical Engineering, Tallin, Estonia, pp. 375-376, ISSN 01400118

[6] Cifrek, M., Tonkovic, S. & Medved, V. (2000). Measurement and Analysis of Surface Myoelectric Signals during Fatigued Cyclic Dynamic Contractions. Measurement, Vol. 27, No. 2, pp. 85-92, ISSN 0263-2241

[7] Cifrek, M., Medved, V., Tonković, S. & Ostojić, S. (2009). Surface EMG Based Muscle Fatigue Evaluation in Biomechanics. Clinical Biomechanics, Vol. 24, No. 4, pp. 327-340, ISSN 0268-0033

[8] De Luca, C.J. (1979). Physiology and Mathematics of Myoelectric Signals. IEEE Transactions on Biomedical Engineering, Vol. 26, No. 6, 313-326, ISSN 0018-9294

[9] De Luca, C.J. (1984). Myoelectrical Manifestations of Localized Muscular Fatigue in Humans. CRC Critical Reviews in Biomedical Engineering, Vol. 11, 251-279, ISSN 0278-940X

[10] De Luca, C.J. (1997). The Use of Surface Electromyography in Biomechanics. Journal of Applied Biomechanics, Vol. 13, 135-163, ISSN 1065-8483

[11] De Luca, C.J. & Knaflitz, M. (1992). Surface Electromyography: What's New?, CLUT, Torino, Italy

[12] Delp, S.L., Loan, J.P., Hoy, M.G., Zajac, F.E., Topp, E.L. & Rosen, J.M. (1990). An Interactive, Graphic-based Model of the Lower Extremity to Study Orthopaedic Surgical Procedures. IEEE Transactions on Biomedical Engineering, Vol. 37, 757-766, ISSN 0018-9294

[13] Delp, S.L., Arnold, A.S. & Piazza, S.J. (2000). Clinical Applications of Musculoskeletal Models in Orthopedics and Rehabilitation. In: Biomechanics and Neural Control of Posture and Movement, J.M. Winters & P.E. Crago, (Eds.), 477-489, Springer-Verlag, ISBN 0387949747, New York-Berlin-Heidelberg

[14] Gandy, M., Johnson, S.W., Lynn, P.A., Reed, G.A.L. & Miller, S. (1980). Acquisition and Analysis of Electromyographic Data Associated with Dynamic Movements of the Arm. Medical & Biological Engineering & Computing, Vol. 18, No. 1, pp. 57-64, ISSN 0140-0118

[15] Farina, D., Fattorini, L., Felici, F. & Filligoi, G. (2002). Nonlinear Surface EMG Analysis to detect Changes of Motor Unit Conduction Velocity and Synchronization. Journal of Applied Physiology, Vol. 93, No. 5, pp. 1753-1763, ISSN 8750-7587

[16] Farina, D., Merletti, R. & Disselhorst-Klug, C. (2004). Multi-Channel Techniques for Information Extraction from the Surface EMG. In: Electromyography – Physiology, Engineering, and Noninvasive Applications, 1 edn, R. Merletti & P. Parker (Eds.), John Wiley & Sons, Inc., ISBN -471-67580-6, Hogoken, New Jersey, pp. 169-203

[17] Frigo, C. & Crenna, P. (2009). Multichannel SEMG in Clinical Gait Analysis: A Review and State-of-the-art. Clinical Biomechanics, Vol. 24, No. 3, pp. 236-245, ISSN 0268-0033

[18] Hermens, J., Freriks, B., Merletti, R., Stegman, D., Blok, J., Rau, G., Disselhorst-Klug, C. & Hägg, G. (1999). SENIAM 8: European Recommendations for Surface Electromyography, Roessingh Research and Development b.v., ISBN 90-75452-15-2, The Netherlands

[19] Konrad, P. (2005). The ABC of EMG. A Practical Introduction to Kinesiological Electromyography, Version 1.0 April 2005, Noraxon INC. USA

[20] Loeb, G.E. & Gans, C. (1986). Electromyography for Experimentalists, The University of Chicago Press, ISBN 0226490149, Chicago & London

[21] Lindström, L. (1973). A Model describing the Power Spectrum of Myoelectric Signals. Part I: Single Fiber Signal. Chalmers University of Technology, Göteborg

[22] Lynn, P.A., Bettles, N. D., Hughes, A.D. & Johnson, S.W. (1978) Influences of Electrode Geometry on Bipolar Recordings of the Surface Electromyogram. Medical & Biological Engineering & Computing, 16, 651-660, ISSN 0140-0118

[23] McMahon, T.A. (1984). Muscles, Reflexes, and Locomotion, Princeton University Press, ISBN 0691083223, New Yersey, USA

[24] Medved, V. (2001). Measurement of Human Locomotion, CRC Press, ISBN 0-8493-7675-0, Boca Raton, Fl., USA

[25] Medved, V. (2007). From Research to Teaching Human Kinesiological Biomechanics: A Zagreb Experience. Challenges in Remote Sensing. Proceedings of the 3rd WSEAS International Conference on Remote Sensing (REMOTE 07), pp. 43-46, ISBN 978-960-6766-17-6 ISSN: 1790-5117, Venice, Italy, November 21-23, 2007 (V. Zanchi, R. Revetria, A. Cecchi, V. Mladenov & A. Zemliak (Eds.), WSEAS Press

[26] Medved, V. & Tonković, S. (1991). Method to Evaluate The Skill Level in Fast Locomotion Through Myoelectric and Kinetic Signal Analysis, Medical & Biological Engineering & Computing, Vol. 29, No. 4, pp. 406-412, ISSN 0140-0118

[27] Medved, V., Tonković, S. & Cifrek, M. (1995). Simple Neuro-mechanical Measure of the Locomotor Skill: An Example of Backward Somersault. Medical Progress through Technology, Vol. 21, No. 2, pp. 77-84, ISSN 0047-6552

[28] Merletti, R., Botter, A., Troiano, A., Merlo, E. & Minetto, M.A. (2009). Technology and Instrumentation for Detection and Conditioning of the Surface Electromyographic Signal: State of the Art. Clinical Biomechanics, Vol. 24, No. 4, pp. 327-340, ISSN 0268-0033

[29] Merletti, R., Farina, D. & Gazzoni, M. (2003). The Linear Electrode Array: A Useful Tool with Many Applications, Journal of Electromyography and Kinesiology, Vol. 13, No. 1, pp. 37-47. ISSN 1050-6411

[30] Merletti, R. & Lo Conte, L.R. (1997). Surface EMG Signal Processing during Isometric Contractions. Journal of Electromyography and Kinesiology, Vol. 7, No. 4, pp. 241-250, ISSN 1050-6411

[31] Merletti, R. & Parker, P.A. (Eds.) (2004). Electromyography – Physiology, Engineering, and Noninvasive Applications, IEEE Press, John Wiley & Sons, ISBN 0-471-67580-6, Hoboken, New Jersey, USA

[32] Nilsson, J., Thorstensson, A. & Halbertsma, J. (1985). Changes in Leg Movements and Muscle Activity with Speed of Locomotion and Mode of Progression in Humans. Acta Physiologica Scandinavica, Vol. 123, pp. 457-475, ISSN 0001-6772

[33] Rose, J. & Gamble, J.G. (Eds.) (2006). Human Walking. Third Edition, Lippincot Williams & Wilkins, ISBN 0781759544, Philadelphia, Pa., USA

[34] Schwartz, M. (1975). Signal Processing: Discrete Spectral Analysis, Detection, and Estimation, Mc Graw Hill, Inc., ISBN 0-07-055662-8, New York, USA

[35] Spiegel., M.R. (1992). Theory and Problems in Statistics. Schaum's Outline Series. Mc Graw Hill, Inc., ISBN 0070602344, New York, USA

[36] Srhoj-Egekher, V., Cifrek, M. & Medved, V. (2010). The Application of Hilbert-Huang Transform in the Analysis of Muscle Fatigue during Cyclic Dynamic Contractions. Medical & Biological Engineering & Computing, Vol. 49, No. 6, pp. 659-669, ISSN 0140-0118

[37] Viitasalo, J.T., Saukkonen, S. & Komi, P.V. (1980). Reproducibility of Measurements of Selected Neuromuscular Performance Variables in Man. Electromyography & Clinical Neurophysiology, Vol. 20, pp.487-501, ISSN 0301-150X

[38] Waterland, J.C. (1968). Integration of Movement, In: Biomechanics I: 1st International Seminar, J. Wartenweiler, J.E. Jokl & M. Hebbelnick (Eds.), 178-187, S. Karger, Basel, Switzerland

[39] Zajac, F.E. (1989). Muscle and Tendon: Properties, Models, Scaling, and Application to Biomechanics and Motor Control, CRC Critical Reviews in Biomedical Engineering, Vol: 17, pp. 359-411, ISSN 0278-940X

doi: 10.1007/978-3-642-15687-8_18

Intrabody Communication in Biotelemetry*

Željka Lučev Vasić, Igor Krois, Mario Cifrek

University of Zagreb, Faculty of Electrical Engineering and Computing

Biotelemetry is remote monitoring, measuring and recording of a living organism's function, activity or condition. Network of sensor nodes placed on or implanted inside the body of a subject is called Body Area Network (BAN). In this work we will describe the principles of a wireless body area network which uses the human body as a transmission medium, namely intrabody communication (IBC). We will describe the limitations set on the IBC systems, describe dielectric properties of the human body as a transmission medium, specify different ways of transmitting signals through the human body and compare characteristics of the IBC systems found in the literature.

Key words:

Biotelemetry, intrabody communication (IBC), capacitive coupling, galvanic coupling.

1. Introduction

Biotelemetry is the use of telemetry methods for sending signals from a living organism over some distance to the receiver [23]. It is used to remotely observe, measure and document human or animal functions, activities, or locations. It consists of several subfields, like medical and human research biotelemetry, implantable biotelemetry and animal biotelemetry. Medical biotelemetry is used to remotely track physiological functions of patients, like body temperature, heart rate, blood

*The chapter was originally published in the book „Wearable and Autonomous Biomedical Devices and Systems for Smart Environment" (Ž. Lučev, I. Krois, and M. Cifrek, Intrabody communication in biotelemetry, vol. 75 of Lecture notes in electrical engineering, pp. 351–368. Springer Berlin Heidelberg, 2010.) by the publisher Springer. The article has been approved for scholarly and non-commercial use.

pressure, ECG, EEG signals, etc., and even to operate devices such as drug delivery systems and prosthetics. Implantable biotelemetry focuses on the transmitter devices implanted in the human or animal being studied, like cochlear implants or implantable pacemakers. Animal biotelemetry is widely used to conduct research on animal behaviour in their natural environment or on animal migration patterns. In this work we will focus our attention on biotelemetry systems used for monitoring human functions only, i.e. medical and implantable biotelemetry systems.

The main constituents of a biotelemetry system are sensors of physiology functions located on the transmitters, transmission path and receivers. Transmitters with sensors are placed on the surface of or implanted inside the human body. In contrast to the medical biotelemetry, the implantable implies the absence of wires as a transmission medium between a transmitter and a receiver. Still, the wires are impractical for monitoring, since they disturb the patient and the medical personnel. Using different wireless technologies provides better freedom of movements and the mobility of the patient, which is of particular importance in a long-term monitoring, every-day activities of non-ambulatory patients and during the surgeries.

Wearable sensor network placed on the human body is called Body Area Network (BAN) and it must comply with the following demands:

- it must not interfere with the human body functions, therefore sensor nodes placed on the body must be in accordance with the existing regulations and guidelines [22, 28];
- it must provide real-time communication;
- low-power requirements;
- no electromagnetic interference between the sensor nodes;
- the presence of the system must not disturb normal movements of the patient, i.e. the sensor network must be lightweight and small-sized.

Two main concerns in a low-power lightweight wireless system design are the distance over which the signal can be received, and the transmitter design, both of which are related to the power source used by the transmitter. Considering that surface mount technology is used in design, the batteries are usually the largest part of the transmitter package. There is a trade off between freedom of the subject's movement and power of transmitter determined by the battery size and capacity: more powerful transmitters with larger batteries can transmit signals over greater distances, but are more difficult to place on or implant in the human body without severely affecting the subject's behaviour and movements.

2. Regulations

Safety restrictions on the effects of exposure to time-varying electric, magnetic, and electromagnetic fields (EMF) are based on established health effects, as described in international guidelines by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) [22] together with different national considerations [28]. The exposure to EMF results in internal body currents and energy absorption in tissues. Protection against adverse health effects requires that proposed basic restrictions are not exceeded.

Guidelines and legal regulations that set basic limitations on exposure to the electromagnetic fields define two types of population [22, 28]. The occupationally exposed population consists of adults who are generally exposed under known conditions and are trained to be aware of potential risk and to take appropriate precautions. The second group is the general public, which comprises of individuals of all ages and of varying health status, and may include particularly susceptible groups or individuals, like children, women and chronically ill people. Therefore, more stringent exposure restrictions are advised for the general public than for the occupationally exposed population, which we take into account in this discussion.

Depending on the frequency of the field, the physical quantities used to specify the basic restrictions on exposure to EMF are as follows, [22, 28], Fig. 1:

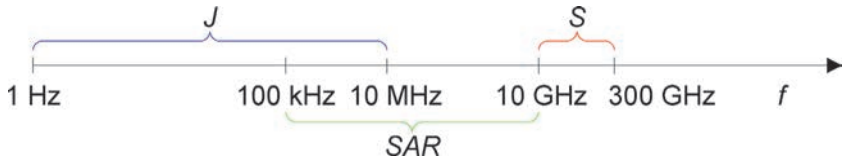


Fig. 1. – Restrictions on exposure to electromagnetic fields.

- current density (J) in the frequency range 1 Hz–10 MHz;
- specific energy absorption rate (SAR) in the frequency range 100 kHz – 10 GHz;
- power density (S), in the frequency range 10–300 GHz.

Restrictions on current density are provided to prevent effects on nervous system functions. The most stringent restrictions are set in the frequency range between 4 Hz and 1 kHz, where the maximum current density is 2 mA/m2, due to a low threshold for nerve stimulation in this frequency range. Below 4 Hz and above 1 kHz the basic restrictions on induced current density increase. Basic restrictions

on the induced current density rms values are given in Table 1., where f is the frequency in hertz.

Table 1 – Maximum current density for head and trunk (rms) [22].

Frequency range, f	J_{max} [mA/m^2]
< 1 Hz	8
1–4 Hz	8/f
4 Hz – 1 kHz	2
1 kHz – 10 MHz	f/500

Between 100 kHz and 10 GHz basic restrictions on specific energy absorption rate are provided to prevent whole-body heat stress and excessive localized tissue heating. Maximum recommended SAR values for the general public population are as follows: a whole-body average SAR 0.08 W/kg, localized SAR in head and trunk 2 W/kg, and localized SAR in limbs 4 W/kg. Restrictions on power density are provided to prevent excessive heating in tissue at or near the body surface. Maximum recommended power density value for the general public is set to 10 W/m2.

A quantity that considers human perception and other indirect effects is contact current. IC is a current that flows when the human body comes into contact with an object at a different electric potential, that is when either the body or the object is charged by an electromagnetic field, [22]. Reference levels for contact current are given in order to avoid shock and burn hazards, for the frequencies up to 110 MHz. The point contact reference levels are shown in Table 2. (f is the frequency in kHz).

Table 2 – Reference levels for time varying contact currents from conductive objects [22].

Frequency range, f	$I_{d,max}$ [mA]
< 2.5 kHz	0.5
2.5–100 kHz	0.2 f
100 kHz – 110 MHz	20

Reference level for current induced in any limb at frequencies between 10 MHz and 110 MHz is below the basic restrictions on localized SAR. This value is set to 45 mA for general public.

3. Wireless biotelemetry systems

In this chapter, we analyse a typical wireless electromyography system as an example of a wireless sensor network. A wireless electromyography system consisting of four EMG sensor nodes placed on the body and a single medical data acquisition system placed a few meters away is depicted in Fig. 2. Sensor nodes and a central unit communicate using a standard wireless link, like WLAN, Bluetooth, RFID or ZigBee. Although often used in biotelemetry, these standards are optimized for other applications: WLAN for data transfer, Bluetooth for voice transmission, RFID for identification and tracking, and ZigBee for industrial applications. Data rate of WLAN and Bluetooth communication modules is fast enough for biomedical applications, but they emit high levels of EMF radiation, which can lead to tissue overheating and irritation. High transmitted power also requires frequent replacement of batteries, which is undesirable in the case of implanted sensors. ZigBee and RFID modules have lower consumption, but their achievable data rates are insufficient for biomedical applications. Common to all of these standards is that they are designed for communication at a distance of several tens of meters, so they inherently generate excessive power.

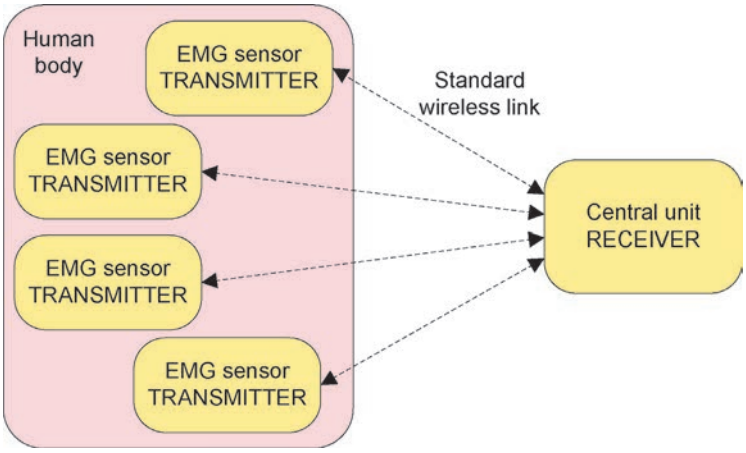


Fig. 2 – Typical wireless electromyography system.

Considering this, a new method of wireless data transmission called intrabody communication (IBC) is developed. IBC uses the human body as a signal transmission medium. It is a short range (up to 2 m) communication with very low consumption (< 1 mW), and a data rate fast enough for biotelemetry applications. All transmitter nodes do not communicate directly with the distant central unit as in Fig. 2, but with a single receiver placed on the human body, in their close vicinity as in Fig. 3, hence further reducing the power consumption. Only the receiver unit communicates with the distant central unit, using a standard wireless protocol. Intrabody communication uses lower frequencies and lower power for short-distance

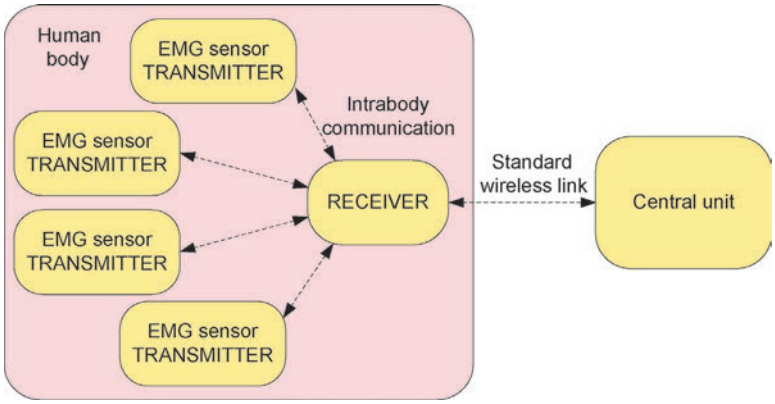


Fig. 3 – IBC wireless electromyography system.

communication through fairly conductive medium of the human body, so the consumption of such a system is lower than of Bluetooth or WLAN systems.

Main components of an intrabody communication system are shown in Fig. 4. IBC system consists of a number of transmitters and a receiver which communicates with the remote central medical unit (due to the low power requirements). Some transmitters can be modified to serve as a signal router. A typical transmitter incorporates sensors of physiological functions for obtaining biomedical signals, signal encoder, a modulator and a coupler. The receiver consists of a detector, demodulator, signal decoder and a signal processing unit.

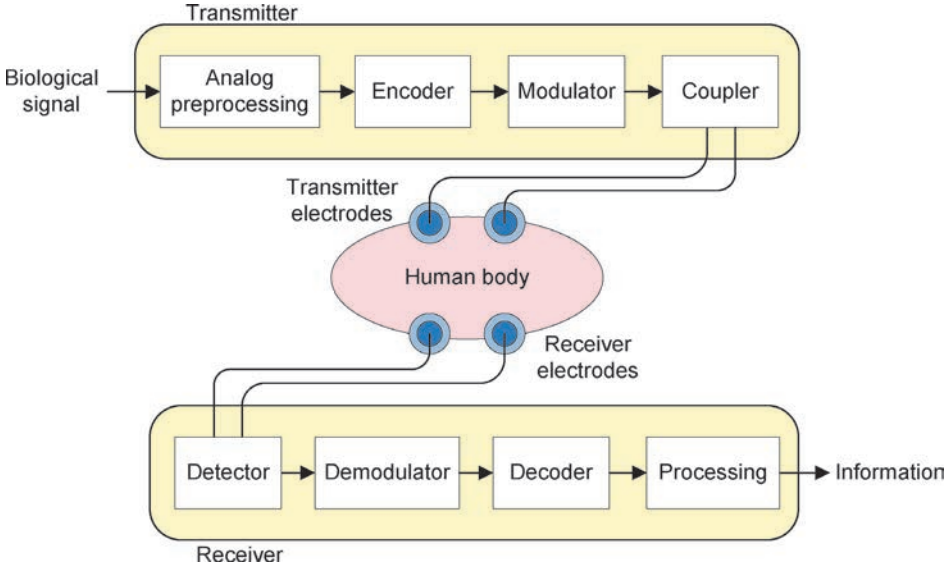


Fig. 4 – Intrabody communication system.

The choice of the modulation, a coupling method, the position of the transmitters with respect to the receiver, and arrangement of the electrodes depend on the coupling method and the application.

4. Dielectric properties of human tissues

The dielectric properties of the human body, electrical conductivity and relative permittivity, determine the flow of electric current and the magnitude of polarization effects, respectively. The dielectric properties of biological tissues for frequency range from 10 Hz to 10 GHz were described and summarized in [9–11] by Gabriel et al. It is shown that the dielectric properties of tissues depend on the type of tissue, frequency, temperature, and the amount of water in a particular tissue, [11]. Frequency dependence of conductivity σ and relative permeability ε_r of human tissues that contribute to the transfer of signals the most (wet and dry skin, muscle in the longitudinal direction, fat and hard bone) at a temperature of 37 °C are shown in Figs. 5 and 6, respectively. It is assumed that

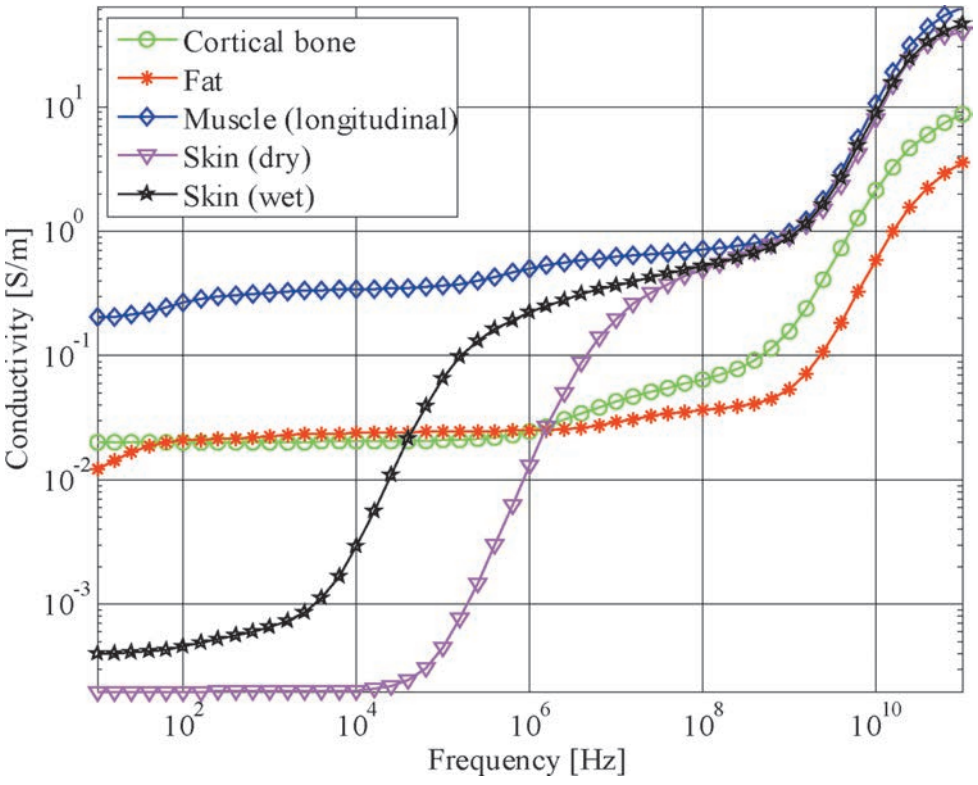


Fig. 5. – Frequency dependence of conductivities for different human body tissues. Adapted from [20].

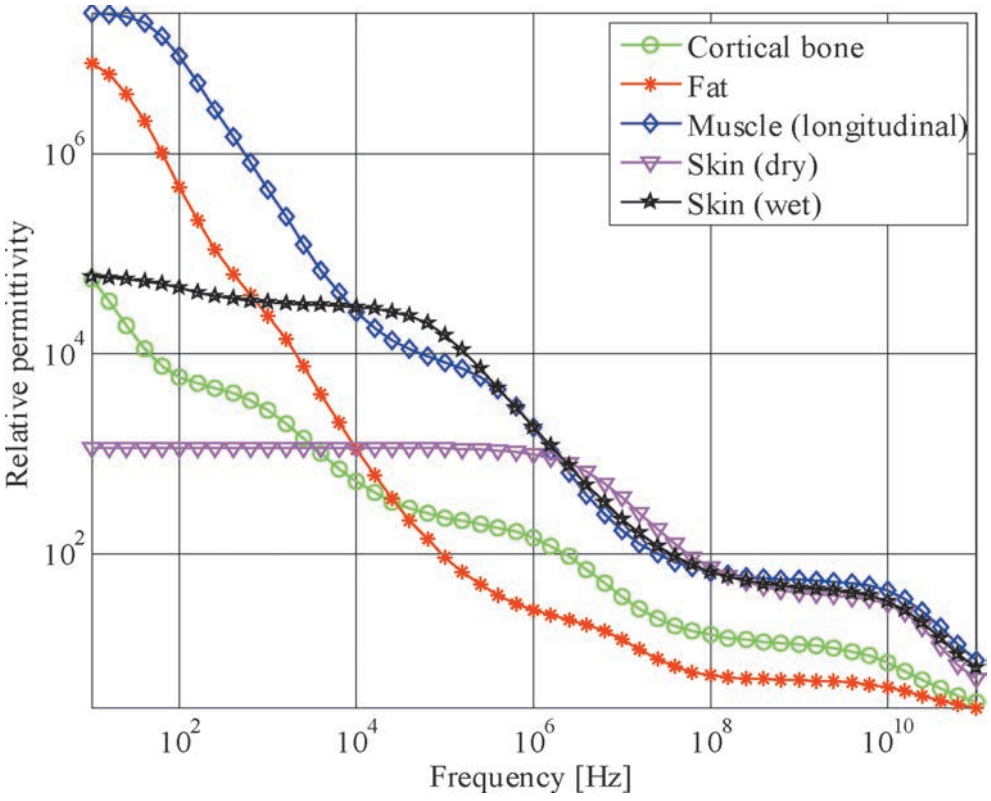


Fig. 6 – Frequency dependence of relative permittivities for different human body tissues. Adapted from [20].

the signals travel through a particular tissue homogeneously, because the penetration depth at the observed frequencies is larger than the dimensions of the body [20].

The signals propagate best through the muscles due to their high conductivity in the whole frequency range, Fig 5. The conductivity and permittivity of the wet and dry skin differ for the frequencies up to 1 MHz. For higher frequencies their values are almost equal to those of the muscles. Above 100 MHz the conductivity of all considered tissues increases rapidly.

The selection of the appropriate carrier frequency for intrabody communication is a trade-off between several requirements: constraints imposed by the safety regulations in order to prevent interference with normal biological signals, ultra-low power requirements, and high tissue conductance at the signal carrier frequency (achieved at higher frequencies).

5. Human body as a signal transmission medium

Transmission characteristic of an IBC system depend on the properties of tissue and a signal path, which is defined by the position of the transmitter relative to the receiver, signal transmission method, signal amplitude, and carrier frequency and type of modulation used. Electric fields external to the body induce a surface charge on the body, which results in induced currents in the body. The distribution of these currents depends on exposure conditions, on the size and shape of the body, and on the body's position in the field. There are different ways to send and receive signals through the human body, which can be reduced to two main methods:

- capacitive signal coupling;
- galvanic signal coupling.

In capacitive signal coupling the signal is controlled by an electric potential between the transmitter electrodes. In galvanic signal coupling the signal is controlled by a current flow through the body.

5.1 Capacitive signal coupling

The use of the human body as a signal transmission medium was proposed at MIT in 1995 [43], and first used in a system known as the Personal Area Network (PAN) [43, 44]. PAN system consists of a transmitter and a receiver, each of which is battery powered and incorporates a pair of electrodes: a signal electrode attached to the human body, and a ground electrode oriented towards the environment. Zimmerman et al. noticed that in the presence of a weak electric field the human body acts as a signal guide and couples the signal electrostatically (capacitively), while the return path is provided by the environment (air and various other objects in the vicinity of the body). Signal transmission is achieved by modulating the voltage between signal electrodes, which is detected and decoded by the receiver. The induced current conducted through the body is in order of magnitude of picoamperes, and is not harmful to the organism.

In the capacitive intrabody communication, as shown in Fig. 7, the induced electric fields are established between all parts of the system that are at different potentials. The transmitter imposes an oscillating potential on the body, relative to the earth ground, causing electric fields E_a between transmitter signal electrode and the body, E_b between the body and the transmitter ground electrode, fields E_c and E_d between the body and the environment, and a field E_s between the body and the receiver signal electrode:

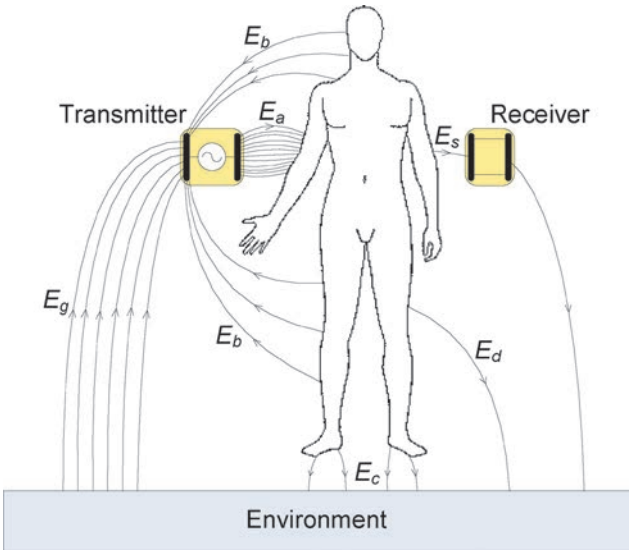


Fig. 7 – Electric fields induced by capacitive intrabody communication.

$$E_a = E_b + E_c + E_d + E_s \tag{1}$$

The return path to the transmitter is closed by the field E_g between the environment and the transmitter ground electrode. Due to the existence of electric fields E_b, E_c and E_d a large part of the transmitted signal closes back to the transmitter ground electrode and the received field E_s is extremely small, but sufficient to obtain the desired information. Since the field E_s additionally decreases with the cube of the distance [44], capacitive intrabody communication can be achieved only at a close distance.

In the capacitive coupling approach the received signal level is affected by the orientation of the transmitter with respect to the receiver, the number of ground electrodes connected to the body, the size of the receiver ground plane, and the surrounding environment ([4], [18], [25]). There is still no agreement on which electrode arrangement is the best, since it highly depends on the application the system is designed for and the signal carrier frequency used. The most common arrangements that yield the largest received signal level are depicted in Fig. 8. In the left one both transmitter electrodes are connected to the body along the line of the signal propagation, and in the right one only the transmitter signal electrode is connected to the body. In both cases the ground electrode of the receiver remains disconnected from the body. The fact that all four electrodes do not necessarily need to be in direct contact with the body allows a successful communication even through the clothes of the subject. Also, increasing the size of the receiver ground plane can strengthen the received signal level, [26].

Fig. 8. – Positions of the electrodes: longitudinal direction (left), two-electrode arrangement (right).

TX is a transmitter, RX is a receiver; S denotes a signal electrode and G denotes a ground electrode, [25].

The signal transmission path highly depends on the configuration of the environment, which must be taken into account when using capacitive IBC approach. Selection of the appropriate carrier frequency is also a very important factor in the design of the IBC system, because an increase of the signal frequency increases the amount of field emitted through the air by radiation, and reduces the field between the receiver and body that we measure.

5.2 Galvanic signal coupling

In the galvanic intrabody communication both transmitter and receiver electrodes are connected to the human body and an alternating current signal is used as a signal carrier, as in Fig. 9. The main (primary) current flow is applied differentially between the transmitter electrodes, while a small part of the current propagates through the body and causes an alternating potential difference between the receiver electrodes. In galvanic coupling approach the transmission path is closed entirely through the subject's body, so the characteristics of a received signal do not depend on the characteristics of the environment, as in the capacitive coupling approach. The main parts of the galvanic IBC system are the same as in the capacitive IBC system, with the exception of the coupler which is adapted to current controlled signal transmission.

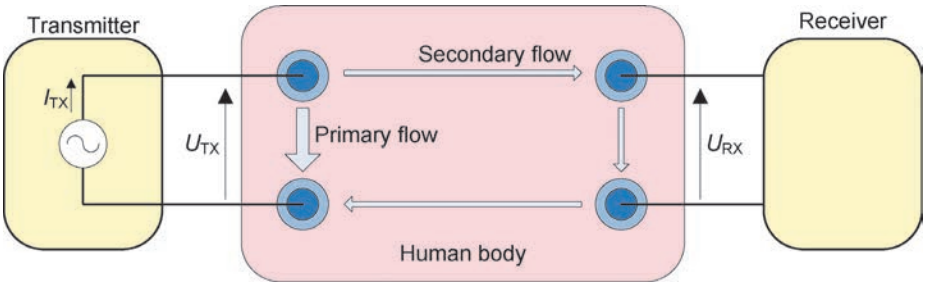


Fig. 9 – Galvanic intrabody communication.

6. IBC technology overview

Several groups from all parts of the world are involved in the development of IBC devices. Various developed IBC systems can be found in the literature, some of which are more oriented to medical applications, while the others are oriented to various non-medical purposes. Since IBC is a new technology, IBC devices are not yet available on the market, but many of the concepts of their application have been thoroughly worked out. The developed systems differ by the coupling method, the coupling amplitude, the chosen frequency range, the signal modulation method, and the achieved data rates.

The concept of Personal Area Networks (PANs) was presented in 1995 to demonstrate how electronic devices on and near the human body can exchange digital information through near-field electrostatic coupling [43]. The first successful PAN prototype used a capacitive signal coupling, a signal with an amplitude of 30 V, 330 kHz carrier frequency, and on-off keying (OOK) modulation. Data rate of 2400 bps and a power consumption of 1.5 mW [44] were achieved. For defining the shape and size of the IBC devices Zimmerman suggested items that are often used in everyday life: watches, credit cards, eyeglasses, identification badges, belts or shoe pads. Depending on the shape and position of an IBC device on the body he suggested a possible application of a new technology: devices such as head-sets, hearing aids, microphones and various indicators may be worn on the head, and identification devices, such as electronic tags, can be placed in a pocket. Indicators, microphones, cameras and speakers can be easily incorporated in a watch, and the heavier devices (PDA, mobile phones, keyboards ...) can be placed around the waist. Also, different physiological sensors can be incorporated in IBC devices and they can monitor the physiological functions of the organism, such as the heart rate, blood pressure, breathing, etc. At the end, Zimmerman suggests the possibility of using devices that are inserted into the shoe, such as pads, which gain the energy from environment, and are used for communication with IBC devices in their vicinity or as transceivers which detect the position and the identity of the people who wear it.

After Zimmerman, at MIT Gray explores the physical limits of intrabody communication [12]. Analysis of the transmission channel showed that the amplifier noise and crosstalk with other IBC devices on the body have the most influence on the received signal. He tested the on-off keying (OOK) and phase shift keying (PSK) modulation at data rates of up to several thousand bits per second. The final version of the hardware [33] was based on frequency shift keying (FSK) modulation and it achieved a data rate of 9600 bps. Analog demodulation at the receiver side was performed using a PLL (phase-locked loop) unit.

Wireless system with very small consumption designed for monitoring the ECG signal was developed by Handa et al. at Waseda University (Japan) [19]. The system consisted of ECG detector placed on the chest of the patient and a receiver placed around the wrist. Alternating current signal with an amplitude of 20 μA was galvanically transmitted between the ECG detector and a receiver. Pulse-width modulation (PWM) with a carrier frequency of 70 kHz was used. Total power consumption of the system was very small, around 8 μW.

M. Fukumoto et al. from NTT Human Interface Laboratories developed a wireless wearable system for the finger-tip typing detection called FingeRing [8]. A test subject on each finger has a ring with a transmitter which registers finger movements when playing or typing. The receiver is located on the wrist, integrated into a wristwatch. Communication system, based on a variant of capacitive IBC, uses an analog frequency modulation with carrier frequencies between 50 kHz and 90 kHz and a consumption of 1.75 mW.

The goal of Derek P. Lindsey et al. [24] was the reduction of volume of implantable de-vices for *in vivo* biometric measurements. They developed a method that uses ionic properties of the human body for the signal transmission, and tested it by measuring tension in anterior cruciate ligament (ACL) grafts of the cadaver. Transmitter platinum electrodes were implanted into the lateral femoral epicondyle, and the signal was detected by surface electromyography electrodes. They studied the effect of transmission frequency, the current injected, the distance between the electrodes, the distance of the electrodes from the joint line and the position of surface electrodes on the signal attenuation. For this application the best results were obtained using a current with amplitude of 3 mA and a frequency modulated (FM) signal with carrier frequency of 37 kHz.

Based on the Zimmerman PAN prototype, K. Partridge et al. from the University of Washington have developed and described a system [31] with which they achieved a data rate of up to 38.4 kbps. They used carrier frequencies of 180 kHz and 140 kHz with FSK modulation and a signal with an amplitude of 22 V. They compared the data error rates and the received signal strengths for different distances between the body and the electrodes, positions of electrodes on the body, size and shape of the receiver signal electrode, and some other factors. They concluded that the size and the shape of the receiver electrodes have only a small effect on signal transmission, in contrast to the distance between the electrodes and the body, which significantly affects the transfer. They also found that the path of the return signal, which closes through the environment, has the greatest influence on the signal transmission, because the useful capacitance of air couplers through the environment is extremely small: it has the order of magnitude in femtofarads [43].

At the university in Chiba (Japan) K. Fujii and K. Ito investigated transmission characteristics of the human body in the IBC system. In [2] they described the FDTD (finite-difference time-domain) simulation of electric field distribution around the numerical model of the human arm approximated by parallelepiped with the following characteristics: dimensions 5 cm x 5 cm x 45 cm, relative permittivity ε_r = 81 and conductivity σ = 0.62 S/m (dielectric parameters of the muscle). IBC transmitter that generates a signal amplitude of 3 V and a frequency of 10 MH was used as a source of an electric field. They showed that the field was formed around the arm and that the received signal attenuation was lower when both the signal and the ground electrode of the transmitter were in contact with the body than when the ground electrode was disconnected from the body. In [3] they added the receiver and monitored the influence of the position of the receiver and transmitter electrodes on the signal transmission through the biological tissue-equivalent solid phantom arm (the same dimensions as the numerical model of the arm) and in FDTD simulation. The configuration in which both transmitter electrodes were attached to the body, while the receiver signal electrode was connected, and the receiver ground electrode disconnected from the body proved to be the best. In [4] they showed that the transmitter electrodes should be placed along the line of the signal propagation from transmitter to the receiver, and that the size of the transmitter and its electrodes significantly affects the strength of the received signal. On the other hand, the change in carrier frequency in interval between 10 MHz and 100 MHz had little effect on signal transmission through the human body. Fujii et al. combined all the previous results in [5], where they compared the results obtained in simulations with the results measured using the biological tissue-equivalent solid phantom arm. They found that the signal was spread as a surface electromagnetic wave along the surface of the skin. Simulation results were confirmed on realistic models of Japanese adults (male and female) [6, 7] and in [7] they presented the spatial distribution of electric field around the numerical model of the arm on which the IBC system was placed.

K. Hachisuka et al. used capacitive method of signal transmission, which they called the electromagnetic wave, for intrabody communication. They showed that in the frequency range from 0.5 MHz to 50 MHz, frequencies around 10 MHz are optimal for the carrier frequency of IBC system with minimal power consumption [15, 16]. The carrier frequency they have chosen for their system was 10.7 MHz, due to the large number of (cheap) components on the market that support it. Measurements were conducted on the phantom hand (described in [15] and [17]) and *in vivo*. The developed transmitters and receivers were the size of 30 mm x 30 mm, weighted around 5 g, and used a 3 V battery as the power supply. The signal amplitude of 3 V and frequency of 10.7 MHz was successfully detected and demodulated at the receiver, for both the frequency (FM) [16] and the digital FSK modulation [15]. Data rate achieved using the FSK modulation was 9.6 kbps. They investigated the optimal position of the transmitter and receiver electrodes for dif-

ferent distances, frequencies and body positions. The two-electrode arrangement, in which only the signal electrode is attached to the body, was found to be 20 dB better in the kilohertz band than the four-electrode arrangement in which all four electrodes are in contact with the body [18]. They suggested that the IBC devices could be used in healthcare, in electronic money transactions, for exchange of business data and for music files sharing.

Japanese phone company NTT (Nippon Telegraph and Telephone Corporation) and its subsidiary NTT DoCoMo Inc. are among the first to use intrabody communication technology to successfully realize communication between electronic devices in everyday life. Their implementation of intrabody communication is called RedTacton [29], and it achieves the data rate of up to 10 Mbps using the capacitive signal transmission method. High data rate is achieved using the electrooptic crystals in the receiver, where change in the electric field with frequency of 10 MHz (which carries the information about the signal), causes measurable change of polarization of the laser beam passing through the crystal [39]. Some of the applications of RedTacton technologies that the NTT proposes are the exchange of data between different devices, identification of users of different services, the use in security systems, etc.

At Waseda University (Tokyo, Japan) a broadband IBC system was developed [34]. The largest attenuation of the received signal was achieved when the signal electrodes of the transmitter and receiver only, having 2 cm in diameter, were connected to the body of the test subjects. Researchers investigated the properties of the human body as a signal transmission medium up to the frequency of 2.5 GHz. Regarding the signal attenuation and time delay, they discovered that BPSK or MSK modulation [35, 36] at the frequencies between 200 MHz and 600 MHz [35] showed the best transmission characteristics for fast broadband intrabody communication. Some of the developed IBC devices [37] were specialized for helping people with special needs, such as a system to help blind and visually impaired people for the orientation in space [37] or a system for mute people to communicate using a wireless keyboard [37]. This system is able to identify which finger the person is using to touch the device and, depending on the selected settings, to reproduce the voice or words on the built-in speakers. They also proposed the use of IBC devices for advertising in trains [37], as well as a perso nalized system for monitoring patient's health [0].

As a part of the European IMEC international corporation with headquarters in Belgium within the Human++ project [21], a sensor network for monitoring vital functions of the human body was developed. Miniature, autonomous and intelligent sensor nodes [13] with extremely low consumption (21 μA @ 3 V) that record the biological signals of the patients (EEG, ECG, EMG – [14]) and send them to different medical central units, where the data is further processed, have been devel-

oped as a part of the research. A very small amount of electrical energy is needed for powering the sensors, and it is obtained by thermal harvesting [32]. The ultra-wide frequency modulation (UWB) between 3 GHz and 5 GHz, with a width of signal spectrum from 500 MHz to 2 GHz is used for communication.

Marc Wegmüller from the ETH Zurich used galvanic intrabody communication [32, 33]. Developed transmitters were connected to the body over two electrodes which generate a current signal with maximal amplitude of 1 mA. Wegmüller investigated the characteristics of the intrabody signal transmission for the signals with frequencies from 10 kHz to 1 MHz. The data rate of 64 kbps was achieved for a carrier frequency of 256 kHz and the QPSK modulation, with a total power consumption of 20 mW. The theoretical results were confirmed by measuring the ECG signal on 20 subjects. It was discovered that the signal transmission depends on the proportion of individual tissues to total body weight and the condition of certain tissues, and that the received signal attenuation was smaller for the greater skin moisture and higher proportion of water in a tissue. The ideal patient for the transmission of signal of frequency from 100 kHz to 500 kHz should have the muscle mass over 45% and the intercellular body water value greater than 60%. Also, a small skin fold thickness positively influences the communication.

At the University of Zagreb, Lučev et al. have used a capacitive IBC approach in designing an electromyography system. The transmitters generate signals with an amplitude of 1.15 V_{pp}, and approximately 1 MHz carrier frequency [25]. Typical achieved data rate was 8 kbps, and the maximal tested data rate was 64 kbps. The researchers investigated relation between the received signal strength and mutual position of the receiver and transmitter in the developed IBC system. *In vivo* measurements showed that the signal propagates through the human body, and not through the surrounding air [25]. The researchers concluded that the received signal strength was higher for longitudinal direction of the transmitter electrodes (Fig. 8., left) for the transmitter located on the same arm as the receiver. For the transmitter located on the opposite arm, the signal was stronger in the two-electrode arrangement (Fig. 8., right) [25, 26]. They also found that increasing the surface reference electrode receiver up to 600 cm^2 improves the transmission of signals [26]. In addition to this, *in vivo* measurements of surface electromyography signals showed that the developed IBC system can be used to monitor muscle fatigue, which is of special importance in kinesiology and rehabilitation [27].

Besides the above applications, it is interesting to mention the use of IBC capsules as miniature endoscopic devices [42]. The capsule that a patient swallows is used to collect data of the state of the desired part of the body from the inside (digestive system) that is then sent through the human tissue to the receiver on the surface of the body.

Table 3 – Comparison of intrabody communication systems. Adopted from [41].

Authors	Year	Coupling method	Coupling amplitude	Carrier frequency	Modulation	Data rate [bps]
Zimmerman	1995	capacitive	30 V	330 kHz	OOK	2.4 k
Fukumoto	1997	capacitive	21 V	90 kHz	FM	0.1 k
Reynolds	1997	capacitive	10 V	70 kHz	FSK	9.6 k
Partridge	2001	capacitive	22 V	160 kHz	FSK	38.4 k
Fujii/Ito	2002	capacitive	3 V	10 MHz	OOK	/
Hachisuka	2003	capacitive	1 V	10.7 MHz	FSK	9.6 k
NTT/ DoCoMo	2003	capacitive	25 V	10 MHz	OOK	10 M
Ruiz/ Shimamoto	2005	capacitive	/	200–600 MHz	MSK	/
IMEC	2005	capacitive	/	2.4 GHz	UWB	/
Lučev	2009	capacitive	1.15 V	1 MHz	FM	64 kbps
Handa	1997	galvanic	20 μA	70 kHz	PWM	0.9 k
Lindsey	1998	galvanic	3 mA	37 kHz	FM	/
Oberle	2002	galvanic	4 mA	60 kHz	CPFSK	4.8 k
Wegmüller	2007	galvanic	1 mA	256 kHz	BPSK	64 k

A comparison between the most important capacitive and galvanic IBC system found in the literature is given in Table 3. The authors of each system, the amplitude and frequency of signals used, the method of modulation used and the maximum achieved data rate are all provided.

7. Conclusion

In this work we described the principles of intrabody communication (IBC) with emphasis on the IBC systems used in biotelemetry. IBC systems utilize conductive properties of the human body for the transmission of the electrical signals. They work at lower frequencies and lower range than standard wireless systems and accordingly, they have lower power consumption. Due to the reduced power consumption, heating and tissue irritation of the patients are lower, and the battery lifetime is longer. We presented the restrictions set on the IBC systems due to the patient safety requirements taking into account the stringent exposure restrictions advised for the general public, which may include particularly susceptible groups or individuals, like children, women and chronically ill people.

Two main intrabody communication methods are capacitive and galvanic signal coupling. In a capacitive signal coupling, the signal is controlled by an electric potential between the transmitter electrodes, and in a galvanic signal coupling, the signal is controlled by a current flow through the body. The literature we reviewed on the subject is abundant with different choices of modulation techniques, coupling methods, and transmitter-receiver arrangements, which all depend on the coupling method and the application. Possible applications of IBC technology are broad: from biotelemetry and medicine to the communication between various electronic devices and security systems.

There is still a need for a detail investigation of the properties of the human body as a transmission medium for IBC signals in order to define criteria for the selection of a coupling method and the optimization of the position of the electrodes. The choice of the modulation, carrier frequency and coupling amplitude must be carefully performed, with respect to the criteria of consumption, probability of the error, and data rate.

References

[1] Alshehab A, Kobayashi N, Kikuchi R, Ruiz J A, Shimamoto S, Ishibashi H (2008) A study on intra-body communication for personal healthcare monitoring system. 10th International Conference on e-health Networking, Applications and Services (HealthCom 2008), Singapore, pp. 219–220

[2] Fujii K, Ito K, Tajima S (2002) Signal propagation of wearable computer using human body as transmission channel. Proceedings of the International Symposium on Antennas and Propagation ISAP-02, pp. 512–515

[3] Fujii K, Ito K, Tajima S (2003) A study on the receiving signal level in relation with the location of electrodes for wearable devices using human body as a transmission channel. IEEE Antennas and Propagation Society International Symposium, vol. 3, pp. 1071–1074

[4] Fujii K, Ito K (2004) Evaluation of the received signal level in relation to the size and carrier frequencies of the wearable device using human body as a transmission channel. IEEE Antennas and Propagation Society International Symposium, vol. 1, pp. 105–108

[5] Fujii K et al. (2005) Study on the transmission mechanism for wearable devices using the human body as a transmission channel. IEICE Trans. Commun., vol. E88-B, no. 6, pp. 2401–2410

[6] Fujii K, Takahashi M, Ito K (2006) Study on the electromagnetic field distributions of realistic Japanese adult male and female models with a wearable device using the human body as a transmission channel. IEEE Antennas and Propagation Society International Symposium, Albuquerque, NM, USA, vol. 1, pp. 2121–2124

[7] Fujii K , Takahashi M, Ito K (2007) Electric field distributions of wearable devices using the human body as a transmission channel. IEEE Transactions on Antennas and Propagation, St. Petersburg, Russia, vol. 5, no. 7, pp. 2080–2087

[8] Fukumoto M, Tonomura Y (1997) Body coupled FingeRing: Wireless wearable keyboard. CHI'97 Conference on Human Factors in Computing Systems, pp. 147–154

[9] Gabriel C, Gabriel S, Corthout S (1996) The dielectric properties of biological tissues: I. Literature survey. Physics in Medicine and Biology, vol. 41, pp. 2231–2249

[10] Gabriel S, Lau R W, Gabriel C (1996) The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz. Physics in Medicine and Biology, vol. 41, pp. 2251–2269

[11] Gabriel S, Lau R W, Gabriel C (1996) The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues. Physics in Medicine and Biology, vol. 41, pp. 2271–2293

[12] Gray M (1997) Physical limits of intrabody signalling. Bachelor thesis, MIT Media Laboratory, Cambridge, MA, USA

[13] Gyselinckx B, Van Hoof C, Ryckaert J, Yazicioglu R F, Fiorini P, Leonov V (2005) Human++: autonomous wireless sensors for body area networks. Proceedings of the IEEE Custom Integrated Circuits Conference, pp. 13–19

[14] Gyselinckx B et al. (2006) Human++: Emerging Technology for Body Area Networks". IFIP International Conference on Very Large Scale Integration, Nice, France, pp. 175–180

[15] Hachisuka K et al. (2003) Development and performance analysis of an intra-body communication device. 12th International Conference on Transducers, Solid-State Sensors, Actuators and Microsystems, Boston, USA, vol. 2, pp. 1722–1725

[16] Hachisuka K et al. (2003) Development of wearable intra-body communication devices. Sensors and actuators A: Physical, vol. 105, pp. 109–115

[17] Hachisuka K, Takeda T, Terauchi Y, Sasaki K, Hosaka H, Itao K (2005) Intra-body data transmission for the personal area network. Microsystem Technologies, vol. 11, no. 8–10, pp. 1020–1027

[18] Hachisuka K et al. (2006) Simplified circuit modelling and fabrication of intra-body communication devices. Sensors and actuators A: Physical, vol. 130–131, pp. 322–330

[19] Handa T, Shoji S, Ike S, Takeda S, Sekiguchi T (1997) A very low-power consumption wireless ECG monitoring system using body as a signal transmission medium. Proc. of the 1997 International Conference on Solid-State Sensors and Actuators, Chicago, USA, pp.1003–1006

[20] IFAC (2000) Dielectric properties of body tissues calculator. Available online: http://niremf.ifac.cnr.it/tissprop/, December 2009.

[21] IMEC Human++ projekt. Available online: http://www2.imec.be/imec_com/imec_com_homepage.php, December 2009.

[22] INCIRP (1998) Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic, and Electromagnetic Fields (up to 300 GHz). Health Physics, vol. 74, no. 4, pp. 494–522

[23] International Society on Biotelemetry ISOB. Available online: www.biotelemetry.org/, December 2009.

[24] Lindsey D P, McKee E L, Hull M L, Howell S M (1998) A new technique for transmission of signals from implantable transducers. IEEE transactions on biomedical engineering, vol. 45, pp. 614–619

[25] Lučev Ž, Krois I, Cifrek M (2009) A multichannel wireless EMG measurement system based on intrabody communication, XIX IMEKO World Congress, Fundamental and Applied Metrology, Lisbon, Portugal, pp. 1711–1715

[26] Lučev Ž, Krois I, Cifrek M, Džapo H, Tonković S (2009) Effects of Transmitter Position and Receiver Ground Plane on Signal Transmission in Intrabody Wireless EMG Measurement System. Proceedings of XI World Congress on Medical Physics and Biomedical Engineering, Munich, Germany, pp. 887–890

[27] Lučev Ž, Krois I, Cifrek M (2010) Application of wireless intrabody communication system to muscle fatigue monitoring. International Instrumentation and Measurement Conference 2010, Austin, TX, USA, accepted for publication

[28] Ministry of Health (2003) Regulations for protection against electromagnetic fields (Pravilnik o zaštiti od elektromagnetskih polja), no. 3306, *in croatian*

[29] NTT RedTacton (2009) Available online: http://www.redtacton.com/en/index.html

[30] Oberle M (2002) Low power system-on-chip for biomedical application. PhD thesis Diss. ETH No. 14509, ETH Zürich, Suisse (*from* [41])

[31] Partridge K et al. (2001) Empirical measurements of intrabody communication performance under varied physical configurations. UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology, Orlando, Florida, USA, pp. 183–190

[32] Penders J et al. (2008) Human++: From technology to emerging health monitoring concepts. ISSS-MDBS 2008, Hong Kong, pp. 94–98

[33] Post E R, Reynolds M, Gray M, Paradiso J, Gershenfeld N (1997) Intrabody buses for data and power. First International Symposium on Wearable Computers, Cambridge, USA, pp. 52 – 55

[34] Ruiz J A, Shimamoto S (2005) A study on the transmission characteristics of the human body towards broadband intra-body communications. Proceedings of the 9th International Symposium on Consumer Electronics (ISCE 2005), pp. 99–104

[35] Ruiz J A, Xu J, Shimamoto S (2006) Propagation characteristics of intra-body communications for body area networks. IEEE Consumer Communications and Networking Conference (CCNC), vol. 1, pp. 509–513

[36] Ruiz J A, Shimamoto S (2006) Experimental Evaluation of Body Channel Response and Digital Modulation Schemes for Intra-body Communications. IEEE International Conference on Communications, Istanbul, Turkey, vol. 1, pp. 349–354

[37] Ruiz J A, Shimamoto S (2006) Novel communication services based on human body and environment interaction: applications inside trains and applications for handicapped people. IEEE Wireless Communications and Networking Conference (WCNC 2006), Las Vegas, NV, USA, vol. 4, pp. 2240–2245

[38] Shimamoto S, Alsehab A M, Kobayashi N, Dovchinbazar D, Ruiz J A (2007) Future applications of body area communications. 6th International Conference on Information, Communications & Signal Processing, Singapore, pp. 1–5

[39] Shinagawa M, Fukumoto M, Ochiai K, Kyuragi H (2004) A near-field-sensing transceiver for intrabody communication based on the electrooptic effect. IEEE Transaction on Instrumentation and Measurement, vol. 53, pp. 1533–1538

[40] Wegmüller M S, Oberle M, Kuster N, Fichtner W (2006) From dielectrical properties of human tissue to intra-body communications. In Proc. of IFMBE, World Congress on Medical Physics and Biomedical Engineering 2006, Seoul, Korea,vol. 14, br. 2, pp. 613–617

[41] Wegmüller M S (2007) Intra-Body Communication for Biomedical Sensor Networks. PhD thesis, Diss. ETH No. 17323, ETH Zürich, Suisse

[42] World Intellectual Property Organization (2008) Capsule type endoscope, patent WO 2008/065839 A1

[43] Zimmerman T G (1995) Personal Area Networks (PAN): Near-field intra-body communication. Master thesis, MIT Media Laboratory, Cambridge, MA, USA

[44] Zimmerman T G (1996) Personal Area Networks: Near-field intrabody communication. IBM Systems Journal, IBM Corp., vol. 35, pp. 609–617

doi: 10.1016/j.cherd.2011.12.014

Non-ideal FLow in an Annular Photocatalytic Reactor

Franjo Jović, Vanja Kosar, Vesna Tomašić, Zoran Gomzi

University of Zagreb, Faculty of Chemical Engineering and Technology,
Department of Reaction Engineering and Catalysis, Savska cesta 16, Croatia

This study deals with the modelling of non-ideal flow in a tubular photocatalytic reactor with thin layer of TiO_2 photocatalyst. The objective was to analyse different level of mixing in the photoreactor applying basic principles of chemical reaction engineering. For this purpose photocatalytic oxidation of toluene was used as the model reaction. Photocatalytic reactor was operated in two different flow modes: classic type of an annular reactor with basically ideal (plug) flow with some extent of dispersion and annular flow reactor acted as stirred tank reactor with mixing of reaction mixture accomplished by recirculation. A series of experiments with step input disturbance at the entrance of the reactor with different air flow was performed in order to achieve better understanding of the reactor hydrodynamics. Several reactor models are applied, such as one dimensional model of tubular reactor at the steady state conditions, axial dispersion model at non-stationary conditions and the model of the continuous non-stationary stirred tank reactor. Numerical methods necessary for solving model equations and parameter estimation were described.

Key words:
Annular photocatalytic reactor; Toluene oxidation; TiO_2; Reactor models

1. Introduction

Volatile organic compounds (VOCs) represent the important class of pollutants, usually found in the atmosphere of all urban and industrial areas. Photocatalytic

*The article was originally published in the journal "Chemical Engineering Research and Design" 9 0 (2012) 1297-1306 by the publisher Elsevier. The article has been approved for scholarly and non-commercial use. The reprint permission granted by the authors.

oxidation, PCO has become increasingly popular as promising alternative to traditional process for the VOC removal (Ertl et al., 1997; Lasa et al., 2005; Nevers, 1995). The UV based PCO technique for air purification can be implemented even under room temperature and atmospheric pressure. Therefore, it may be more energy efficient than other conventional techniques.

Commercial TiO_2 photocatalysts seems to be well suited for the purification of indoor air. However, there are also some drawbacks associated with their application. This is mostly due to the problems related to deactivation of TiO_2 catalyst after working for a certain period of time (Dezhi et al., 2005; d'Hennezel et al., 1998; Maira et al., 2001; Mendez-Roman and Cardona-Martinez, 1998), as well as due to complex and changeable operating conditions typical for indoor air application. In the past two decades, a lot of investigations have been conducted in this field of research. Catalysts are usually coated on the wall of a photoreactor (thin film reactor) or photoreactor is filled with materials acting as catalyst support. Various supports of TiO_2 are concerned, such as pellets (Bouazza et al., 2008), non-woven fibre textile (Ku et al., 2001), paper holders (Iguchi et al., 2003), metal foam (Ibhadon et al., 2007), zeolite panels (Ichiura et al., 2003) and rashing rings (Quici et al., 2010). Investigation of optimal reactor configurations has also become an important research area in the field of PCO (Lasa and Rosales, 2009). Different designs of laboratory reactors are used, including honeycomb monolith reactor (Du et al., 2008; Nicolella and Rovatti, 1998; Raupp et al., 2001; Huang and Li, 2011.), flat-plate reactor (Salvadó-Estivill et al., 2007a), fixed bed annular reactor (Alberici and Jardim, 1997; Bouzaza et al., 2006; Jeong et al., 2004; Keller et al., 2003), batch reactor (Kim et al., 2002), semi-batch reactor with quartz flat window (Demeestere et al., 2004; Zuo et al., 2006), circulating fluidized bed (CFB) reactor (Dibble and Raupp, 1992; Lim and Kim, 2005; Prieto et al., 2007; Sekiguchi et al., 2008), micro channel reactor (Ge et al., 2005), TiO_2-coated fibre-optic cable reactor (Denny et al., 2009; Peill et al., 1997), annular venturi reactor (Photo-CREC-air) (S. Romero-Vargas Castrillón et al., 2006) and others.

During recent years, various mathematical models describing the flow through photocatalytic reactor, interaction between the light, polluted air and catalytic surface are used and several research strategies are described. Nicolella and Rovatti (1998) presented distributed parameter model for photocatalytic oxidation of air contaminants in monolith reactors. Tomasic et al. (2008) developed and compared one dimensional (1D) and two dimensional (2D) heterogeneous models of an annular photocatalytic reactor based on assumed ideal and laminar flow through the reactor. Imoberdorf et al. (2007) described a proposal for scaling-up of photocatalytic reactors designed as catalytic walls coated with a thin layer of

titanium dioxide. Interesting approach to the modelling of PCO reactors is reported by Zhang et al. (2003). Based on the analogy between heat and mass transfer for

heat exchangers they developed reactor model with two parameters, the fractional conversion and the number of mass transfer units as the main parameters influencing the photo oxidation performance of PCO reactors. In general, simplified hydrodynamic models, such as plug flow or completely mixed flow are proposed and discussed in the literature. Obviously, by using such approach it is impossible to correctly describe the reactor hydrodynamics and to make conclusions about the performance of the photoreactor which can be greatly influenced by the reactor hydrodynamics. As known, the computational fluid dynamics (CFD) is emerging engineering tool which can be used for design and optimization of chemical reactors due to coupling the reactor geometry and reaction mixture flow thorough reactor. Thus, several researchers use CFD to simulate UV-reactor performance through the integration of reactor hydrodynamics, radiation distribution and UV reaction kinetics (Taghipour and Mohseni, 2005; Salvadó-Estivill et al., 2007b; Queffeulou et al., 2010).

Generally, the performance of an annular photocatalytic reactor used for removal of volatile organic compounds from the gas phase can be affected by operation parameters, such as reactor configuration (reactor geometry, the shape of the reactor inlets and outlets with respect to the reactor axis) and reaction conditions (initial concentration of reactant(s), humidity, the light source and intensity, total flow rate of the reaction mixture, etc.). Obviously, the overall performance of an annular reactor is greatly influenced by the reactor hydrodynamics due to the fact that the reacting fluid usually flows through the reactor with various degree of mixing. However, little work has been done in this field, especially with regard to the reactors used for photocatalytic oxidation in the gas phase. As well known, three main characteristics can be used to describe non-ideal reactors: distribution of residence time in the system, the quality of mixing and the model used to describe the system. The objective of this work was to carry out a detailed computational and experimental study of hydrodynamics in the annular photoreactor, to develop appropriate mathematical models as well as to compare the results with experimental measurements.

2. Experimental

Detailed description of the experimental set-up and procedure can be found elsewhere (Tomasic et al., 2008). In short line, high purity synthetic air (20.5% O_2 in N_2, *Messer*) was used as oxidant and carrier gas. The appropriate concentrations of the toluene (*Aldrich*) and water were obtained by the vaporization of the organic compound and water at the specified flow rates of the gas carrier through the saturators. The flow rates were regulated using the mass flow controllers (*Cole Palmer*). The secondary flow of the air was introduced into the mainstream to di-

lute the reaction mixture and finally to obtain desired concentration of the reactant at the reactor inlet. The temperature of the liquid containing saturators (toluene and water) was maintained at 20 °C. All measurements were carried out at the room temperature, atmospheric pressure, at the initial concentration of toluene of $2.68 \cdot 10^{-3}$ g dm^{-3} and the relative humidity RH of 50%. The total flow rate of reaction mixture was in the range from $(41–123) \cdot 10^{-3}$ dm^3 min^{-1}. The membrane peristaltic pump was applied to circulate a particular amount of the fluid containing a reactant to the reactor entrance. Experiments were performed in the range of the recirculation flow rate, v_R from 0.5 to 2.5 dm^3 min^{-1} (corresponding to recirculation ratios, R from 12 to 61). Recirculation ratio was defined as flow rate of a reaction mixture circulated to the reactor entrance and flow rate of a reaction mixture that is in effluent from the reactor system.

2.1 Preparation of the catalyst layer

The commercial TiO_2 powder catalyst was supplied by *Degussa* (*Aeroxide* P-25). Crystal structure of the catalyst was primarily based on anatase (70% anatase and 30% rutile). The catalyst surface area determined using the BET method (Micromeritics Acusorb 2010) was 58.16 m^2 g^{-1}. Details about characterization of TiO_2 catalyst have been reported previously (Tomasic et al., 2008). The inner wall of the reactor was coated with TiO_2 aqueous suspension. The suspension was poured into a rotating tube of reactor and simultaneously dried in the hot air. Coating was repeated several times until desired thickness of the catalytic layer was achieved. The reactor was assembled with the lamp after the catalyst had been attached to the reactor's wall.

2.2 Investigation of the photoreactor performance and hydrodynamics

Two configurations of the annular photoreactor were used. The geometric parameters were almost the same in both cases (Table 1), and difference was in the shape of the reactor inlets with respect to the reactor axis, e.g. in one case the reactor design ensured the axial introduction of the reaction mixture at the entrance of the reactor (reactor I), and in the second case the radial introduction (reactor II) (Fig. 1).

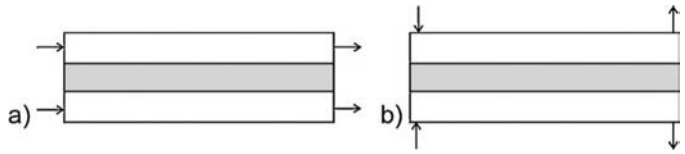


Fig. 1 – Scheme of the photocatalytic annular reactors: (a) reactor with axial, and (b) radial reaction mixture inlet.

Table 1 – The geometry properties of two different configurations of the annular photocatalytic reactor.

Properties	Reactor configurations	
	Reactor I	Reactor II
Reactor length, L/cm	26,5	28,5
Internal diameter of the outer tube, R_i/cm	3,03	3,03
External diameter of the inner tube, R_e/cm	2,50	2,50
Volume, V/cm^3	61,0	65,6
Introduction of the reaction mixture at the reactor entrance	Axial	Radial

The photoreactor consists of two concentric reactor's tubes. The UV lamp was located in the central part of the reactor, protected by the inner tube of the reactor. Inner tube was made of quartz glass because of its transparency for UV light. The TiO_2 layer was coated on the internal glass surface of the outer tube of the reactor. Illumination of the catalyst layer was provided by an 8 W fluorescent black light lamp (*Sylvania*®). The reaction gas mixture flowed between the reactor's inner and outer tube.

Sets of the experimental measurements in the annular tubular reactors with and without recirculation have been carried out to obtain kinetic and hydrodynamic parameters under the following conditions:

- in each series of measurements, the flow and composition of the inlet reaction mixture (toluene and moisture concentration) was kept constant,
- the residence time inside the reactor was changed by varying the percentage of illuminated catalytic surface area as specified (33, 50, 66, 75 and 100%); the percentage of illuminated catalytic surface area was changed by covering the catalyst layer by material impermeable to UV light.

The reaction products were measured on line using a Carbowax 20 M column (60/80 mesh) installed in a gas chromatograph (Shimadzu GC-2014) equipped with a flame ionization detector (250 °C), with nitrogen as the carrier gas.

3. Results and discussion

Our preliminary research has shown that problem of TiO_2 catalyst deactivation can partially avoid by performing the photocatalytic reaction in annular reactor operated in a recirculation mode. Generally, this type of reactor is called a recycle reactor.

However, in this text we use the term annular reactor with recirculation. The possible explanation for higher stability of catalyst in the annular reactor with recirculation is higher flow rate of the reaction mixture passing over the catalyst layer leading to the higher rate of the removal of undesired by-products (benzaldehyde and benzoic acids) responsible for the catalyst deactivation from the surface of the catalyst. This mode of operation might be meaningful in photocatalytic reaction with very large film-diffusion resistance, since the rate of the photocatalytic reaction increases due to a decreasing of the film-diffusion resistance when reactant fluid circulate at high flow rate. Moreover, the behaviour of annular photocatalytic reactor with recirculation differs from those for standard type of an annular reactor (without recirculation) depending on recirculation ratios. According to the theory, at the recirculation ratio, R, higher than 20 (e.g. at $R > 20–25$) it is possible to use model of continuous stirred tank reactor (CSTR) corresponding to ideal (or complete) mixing. In this study we will try to examine some situations where the fluid in a reactor is neither well mixed nor ideal (or plug) flow and experimental RTD are used to evaluate the parameters in the proposed models.

The residence time distributions (RTD) are determined experimentally by using step input of the tracer concentration (in our case toluene) in a system with constant volumetric flow rate and then measuring the tracer concentration at the reactor exit as function of time. As mentioned, experiments are conducted both in the presence and in the absence of photocatalytic reaction under isothermal conditions.

3.1 Flow models for the case without reaction

Based on the physical picture it is possible to assume hydrodynamics in the tubular reactors and derive the appropriate reactor model. Using the experimentally determined RTD it is possible to estimate deviations among the proposed models. Several hydrodynamic flow models can be applied, such as laminar flow model and axial dispersion model. For systems with considerable mixing, the tanks-in series model or gamma model can be also acceptable to describe non-ideal reactors.

3.1.1 Model I – axial dispersion model

The axial dispersion model, Eq. (1) is quite often used to describe flow in non-ideal tubular reactors (Fogler, 1986; Nauman, 2008). This is good approximation for most tubular reactor with small deviations from plug-flow. In this study the dispersion model was checked by RTD obtained from step response measurements. In a series of experiments the RTD curves of the toluene concentration (tracer) are determined at the reactor exit after positive step-function input (at $t = 0$ the constant airflow of toluene from bubblers is introduced in the reactor). Fig. 2 shows re-

sponse to the step-function input (e.g. toluene concentration at the reactor exit versus time). As can be seen, significant dispersion was observed and confirmed by the calculation of the Peclet number, as dispersion model parameter, Eq. (1).

$$\frac{1}{Pe}\frac{\partial^2 c_A}{\partial x^2} - \frac{\partial c_A}{\partial x} = \frac{\partial c_A}{\partial \theta} \tag{1}$$

3.1.1.1. Analytical solution. Analytical solution of Eq. (1) is possible only for some type of the boundary conditions and for other types of boundary conditions only numerical solution of Eq. (1) is possible. The initial conditions corresponding to the positive step input of toluene concentration are:

$$\begin{gathered} \text{for } \theta < 0, \text{ and } x > 0,\ c_A = 0 \text{ and} \\ \text{for } \theta < 0, \text{ and } x = 0,\ c_A = c_{A0} \end{gathered} \tag{2}$$

There are two types of boundary conditions that can be consider to solve Eq. (1), e.g. for closed and open vessels. Boundary conditions in which the analytical solution is possible (corresponding to the physical representation of the so called open – open system) are:

$$\begin{gathered} \text{at } x = 0,\ c_A = c_{A0} \\ \text{at } x = 1,\ c_A = c_A \end{gathered} \tag{3}$$

The analytic solution of Eq. (1), with initial, Eq. (2) and boundary conditions, Eq. (3), is:

$$c_A = \frac{C_{A,0}}{2}\left[1 - \operatorname{erf}\left(\frac{1}{2}\sqrt{Pe}\,\frac{1-\theta}{\sqrt{\theta}}\right)\right] \tag{4}$$

3.1.1.2. Numerical solution. Partial differential equation, Eq. (1) together with corresponding initial and boundary conditions, Eq. (2) and Eq. (3) belongs to the parabolic type of equations. In this study it was solved numerically by using method of lines (Schiesser, 1991).

Peclet number, *Pe* as a model parameter was estimated from the experimental data using the non-linear regression analysis (ID algorithm). Estimated values of the *Pe* numbers were 5.76 for the analytical solution and 4.06 for numerical solution of dispersion model, Eq. (1). Apparently, the values of *Pe* number show significant dispersion for both solutions of dispersion model. This means that the system is far of plug flow conditions and some degree of mixing is present, probably due to the

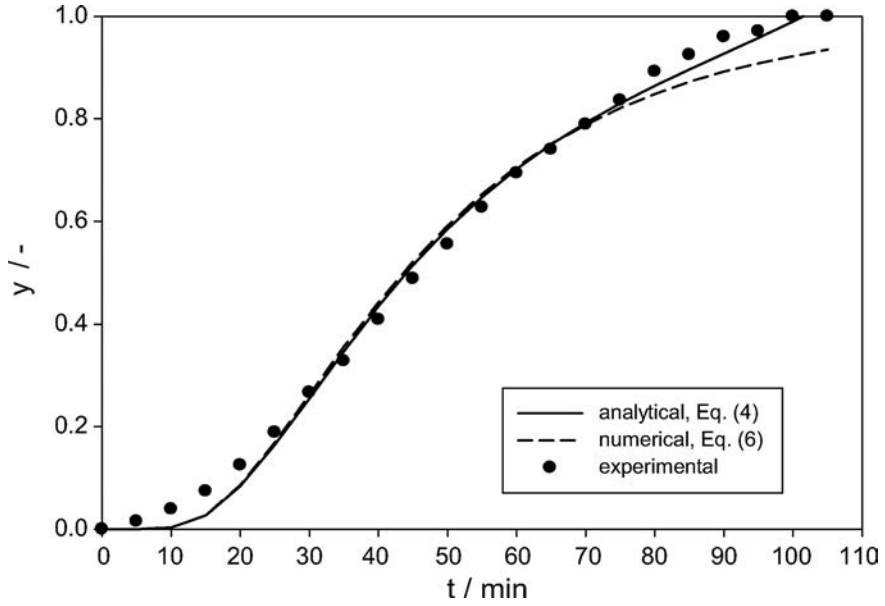


Fig. 2 – Response curves (positive step input) for dispersion model. Points – experimental data, lines analytical solution of the model given by Eq. (4) and numerical solution given by Eq. (6), (reaction condition: $v_0 = 41 \times 10^{-3}$ dm^3 min^{-1}).

flow channelling or misdistribution at the entrance and exit from reactor. As shown in Fig. 2 results based on analytical and numerical solutions are almost the same and show good agreement in mean square deviations (SD = 5.12 and 8.43 × 10^{-3}) between normalized measured and calculated data.

3.1.2 Model II – flow model in the annular reactor with recirculation

Depending on recirculation flow rate, two extreme situations can be observed. When the recycle rate becomes very large, ideal stirred tank performance can be assumed (or ideal mixing) and when the recycle is zero ideal or plug flow operation results. Assuming ideal mixing (or ideal stirred tank reactor) due to recirculation, exit concentration of component A as results of the positive step change in inlet concentration is given by Eq. (5):

$$c_A = c_{A,0}\left(\exp\left[-\frac{t}{\tau_m}\right]\right) \tag{5}$$

or, in dimensionless form,

$$y = 1 - \exp\left[-\frac{t}{\tau_m}\right] \tag{6}$$

Initial conditions corresponding to Eq. (6) are:

$$\begin{aligned} t < 0, \quad y = 0 \\ t \geq 0, \quad y = 1 \end{aligned} \tag{7}$$

Model equation, Eq. (6) together with initial conditions Eq. (7) gives the concentration of tracer (toluene) at any time t. Parameter defined as ratio of the reactor volume, V and volumetric flow rate v_0 is the space time (or holding time), 0 and it should be equal to the mean residence time that tracer spends in the reactor, τ_m if there is no deviation from the ideal reactor performance.

Eq. (6) was applied to the experimental data in Fig. 3 at the constant total flow rate of the reaction mixture and at different recirculation ratios (R = 12–60). Testing of experimental data to the corresponding curve was performed using the integral method. Only adjustable parameter was the mean residence time, τ_m. As can be seen from the data given in Table 2, estimated values of the mean residence time in all experiments deviate from the actual values obtained assuming conditions of ideal mixing. Obviously, the fluid flow through the reactor is the result of the convective and dispersive flow. Taking into account the shape of the curves shown in Fig. 3 it can be concluded that deviations from the ideal flow conditions are greater for smaller values of recirculation ratios. However, for the largest recirculation ra-

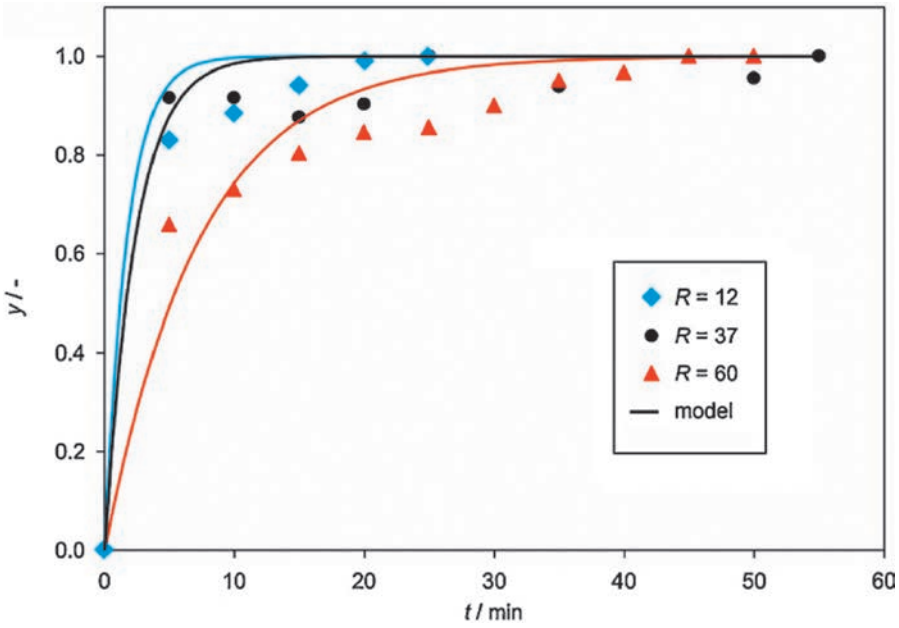


Fig. 3 – Response to the positive step input in reactor with recirculation under the constant flow of v_0 = 41 cm^3 min^{-1} and with different recirculation ratios, R.

Table 2 – Estimated values of the mean residence time as parameter in model, Eq. (6) for experiments carried out in the annular reactor with recirculation and with step tracer input. Fluid flow through the reactor was constant (v_0 = 41 cm^3 min^{-1}).

Recirculation ratio/R	τ_0/min	τ_m/min	SD × 10^2
60	1.60	7.35	2.32
49	1.21	5.21	2.15
37	1.07	3.28	2.26
24	0.80	2.26	1.32
12	0.53	1.72	2.62

tio (R = 60) a quite good agreement was observed indicating that in such case even a simple model of CSTR can be used to describe flow in experimental reactor employed in this work.

3.1.3 Model III – CFD modelling of an annular reactor

Diversified flow in reactor can be expected especially regarding reaction mixture inlets and outlets. Appropriate inlet/outlet design of the reactor can results in an optimum flow distribution while creating sufficient mixing could provide a better reactor performance. The flow pattern in an actual reactor can be directly determined by solving the Navier–Stokes equations. First step in clarifying flow through the reactor is to construct a model which simulates the actual situation. In order to solve and represent the flow pattern, the equations can be solved by different mathematical methods (such as finite elements) using *Fluent software*. In this study the simulation of the system was performed with a three-dimensional (3D), steady state and laminar flow model. The main equations for modelling the system are continuity (8) and momentum (9) equations (Mohseni and Taghipour, 2004; Taghipour and Mohseni, 2005).

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \vec{u}) = 0 \tag{8}$$

$$\frac{\partial(\rho u)}{\partial t} + \nabla(\rho u \vec{u}) = \nabla P + \nabla \times (\bar{\bar{\tau}}) + \rho g \tag{9}$$

Assumption of no compressible flow can be made because flow velocity is much smaller then sound velocity (u/c < 0.3). During development of 3D model it is important to construct mesh of finite elements (volumes) which properly describes real case. It is also necessary to make dense mesh at the entrance of the reactor (and exit from the reactor) where whirl may occur and where large velocity gradient can be present.

The differences between radial and axial way of introducing reaction mixture in an annular reactor are simulated by CFD and results can be seen on Fig. 4. The reactor with the radial inlet of the reaction mixture has a good mixing at the reactor entrance, accompanied by formation of whirl (Fig. 4a). With the axial inlet of the reaction mixture in reactor stream jet penetration is deeper (Fig. 4b), so the working space inside the reactor is decreased, without formation of the uniform velocity profile. This results in loss of reactor efficiency. The radial introducing of the reactants is beneficial, because of the larger working space and higher efficiency inside the reactor.

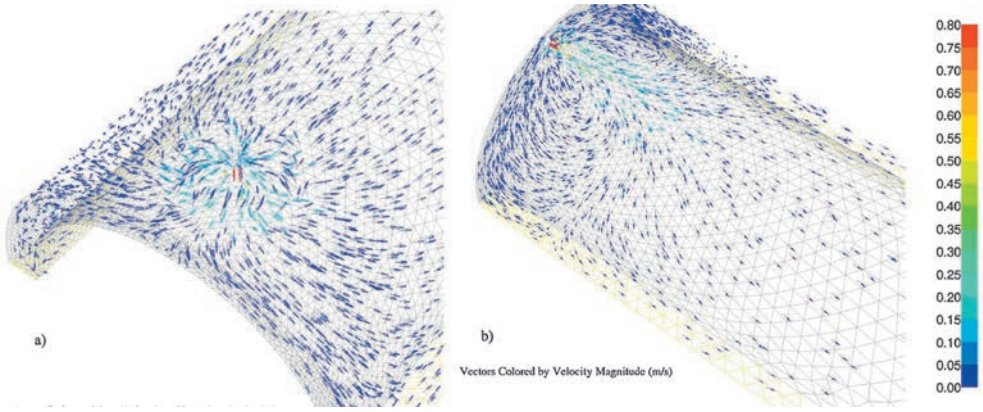


Fig. 4 – Velocity vectors on the reactor entrance: (a) Reactor I, (b) Reactor II.

3.2 Modelling of photocatalytic reactor in the presence of chemical reaction

3.2.1 Model IV – plug-flow tubular reactor at stationary conditions (one – dimensional model)

For the modelling of catalytic reactor different models can be used. Annular photocatalytic reactor operating in the mode without recirculation can be modelled with the simple model assuming *ideal flow* and steady state conditions. Basically, this is possible using the one – dimensional (1D) heterogeneous model. Taking into account the isothermal operation, model is composed only from the mass balance for the reactant both in the fluid and on the surface of the catalyst. Mathematical complexity of the model depends only on the form of kinetic model of the surface reaction rate. Due a very thin catalyst layer on reactor inner wall there is no mass transfer inside the catalyst layer (it can be neglected). Under the influence of light the catalytic reaction takes place only on the surface of deposited TiO_2.According

to the above assumption, the overall mass balances in the solid phase Eq. (10) and in the gas phase (11) constitute the model:

$$r_A = f(c_{A,s}) = k_g \cdot a \cdot (c_{A,s} - c_A) \tag{10}$$

$$u\frac{\mathrm{d}c_A}{\mathrm{d}z} = k_g \cdot a \cdot (c_{A,s} - c_A) \tag{11}$$

The initial conditions are:

$$\text{at } z = 0,\ c_{A,s} = 0,\ c_A = c_{A,0} \tag{12}$$

It is important to mention that for one series of measurements (with the same flow) the residence time was changed by varying length of the reactor, and thus size (volume) of catalyst. Therefore, the variable is the length of the reactor, Eq. (13) or the residence time, Eq. (14) and corresponding expressions are:

$$u\frac{\mathrm{d}c_A}{\mathrm{d}z} = k_g a(z)(c_{A,s} - c_A) \tag{13}$$

$$\frac{\mathrm{d}c_A}{\mathrm{d}\tau} = k_g a(\tau)(c_{A,s} - c_A) \tag{14}$$

The mass transfer coefficient, k_g in Eqs. (13) and (14) can be determined for every flow rate using the empirical correlation:

$$k_g = \frac{Sh \cdot D_A}{2R(1-\kappa)} \tag{15}$$

There are different literature correlations for calculation of Sherwood number. In this work we have used the following (Mohseni and Taghipour, 2004; Taghipour and Mohseni, 2005):

$$Sh = 0.705\left[Re\frac{d}{L}\right]^{0.43} Sc^{0.56} \tag{16}$$

Estimated reactor parameters are the reaction rate constant in the kinetic model given by Eq. (17) based on assumption of the pseudo-first order reaction rate and mass transfer coefficient, Eqs. (13) and (14):

$$-r_A = k_A c_A \tag{17}$$

Parameter estimation is performed using a modified differential method of data analysis and the Nelder–Mead method of non-linear optimization. As can be seen from the results for each series of measurements a satisfactory agreement was obtained (Table 3, Fig. 5). The estimated reaction rate constant, *kA*, is approximately equal for all series of experiments. This agreement is expected because the reaction conditions in all experiments were the same (the catalyst and the composition of the inlet mixture), except the total flow rate of the reaction mixture. However, the influence of the flow rate on the mass transfer coefficient, *kg* was significant. Comparison can be made between the calculated values of *kg* obtained using Eq. (16) and values estimated from the experimental data, Eqs. (13) and (14).

Table 3 – Estimated reactor parameters and agreement with the model given by Eq. (13)–(15).

v_0/cm^3 min^{-1}	k_A/min^{-1}	k_g/cm min^{-1}		SD × 10^2
		Estimated	calc. Eq. (15)	
41	1.04	0.62	1.67	2.82
61.5	1.02	1.21	1.99	0.43
82	1.05	1.02	2.25	4.82
102.5	1.07	1.23	2.48	6.51
123	1.05	1.77	2.68	5.28

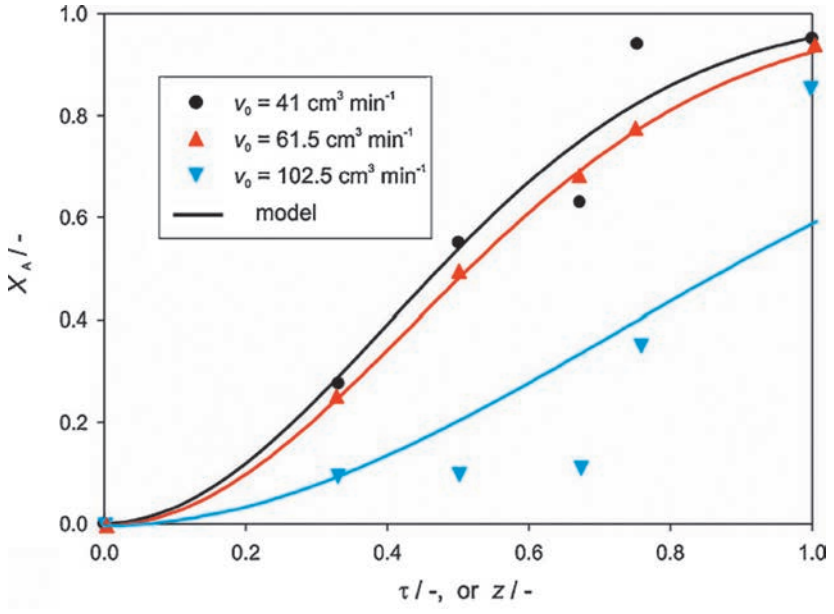


Fig. 5 – Dependence of the toluene conversion vs. dimensionless residence time or reactor length for different total flow rate, v_0.

3.2.2 Model V – non-stationary tubular reactor with reaction under assumption of the axial dispersion flow (one – dimensional model)

As pointed out previously, shapes of curves showing conversion versus time obtained in experiments conducted in annular reactor point out the fact that non-ideal flow is present in our reaction system. In a series of experiments, concentration of toluene was changed from the stationary values to the new one by switching on UV light, giving step input in time $t = 0$ (unsteady-state). Then, the exit concentrations of toluene were measured during the time. Flow of the reaction mixture through the reactor was described by *axial dispersion model* (*Model I*). Assuming the non-steady state, the mass balances for both, the solid phase, Eq. (18) and the gas phase, Eq. (19) in dimensionless form have the following form:

$$k_g a(c_A - c_{A,s}) = -r_A \tag{18}$$

$$\frac{\partial c_A}{\partial \theta} = \frac{1}{Pe}\frac{\partial^2 c_A}{\partial x^2} - \frac{\partial c_A}{\partial x} - \tau k_g a(c_A - c_{A,s}) \tag{19}$$

Chemical reaction on the surface is described by first-order reaction model, Eq. (17). Based on the physical picture of the system, the appropriate boundary and initial conditions for non-stationary operation are:

$$\text{reactor entrance, } x = 0, \begin{cases} \theta < 0, & c_{A,t,x} = 0 \\ \theta \geq 0, & \dfrac{\partial c_{A,\theta,0}}{\partial x} = Pe(c_{A,0} - c_{A,\theta,0}) \end{cases} \tag{20}$$

$$\text{reactor exit, } x = 1, \left| \theta \geq 0, \quad \frac{\partial c_{A,\theta,1}}{\partial x} = 0 \right. \tag{21}$$

$$\text{initial condition, } \left| \theta = 0, \quad c_{A,0,0} = 0 \right. \tag{22}$$

Model equations, given by Eqs. (18)–(22) are solved numerically using method of lines, as mentioned previously. For a series of experiments with various residence times differential analysis was performed aimed to estimate parameter *Pe,* (Table 4). As can be seen relatively small values of *Pe* numbers were obtained. One such test for the two experiments with different residence time is shown in Fig. 6. In fact, only one parameter, *Pe* number was included in parameter estimation, whereas other constants, *kA* and *kg* were taken from *Model IV* (Table 3.). Based on the estimated value of *Pe* number it can be concluded that the dispersion in the reactor is significantly present (Levenspiel, 1999.). It is evident that by further reduction in the length of the reactor (smaller residence time) dispersion is reduced, which is in accordance to the definition of Peclet or dispersion number.

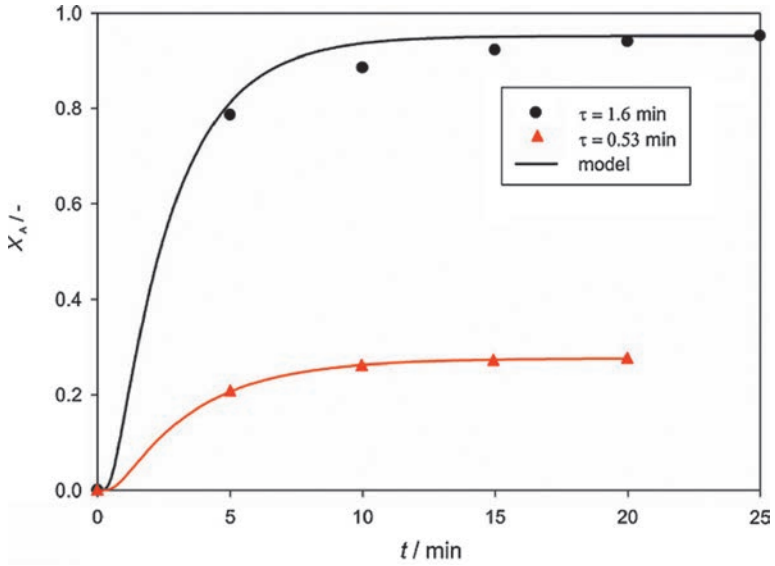


Fig. 6 – Testing the axial dispersion model with chemical reaction (v_0 = 41 cm^3 min^{-1}).

Table 4 – Test results obtained using the dispersion model of annular tubular reactor.

Reactor length, z	τ_0/min	Pe	SD × 10^2
$z = L$	1.60	3.67	0.921
z = 3/4 L	1.20	4.91	0.767
z = 2/3 L	1.07	5.32	5.415
z = 1/3 L	0.53	5.63	0.077

3.2.3 Model VI – continuous non – stationary stirred-tank reactor (CSTR)

In this case, experiments are carried out in annular tubular reactor with different recirculation ratio in non-stationary operation. During the experiment performed as a function of time, the volumetric flow rate and inlet concentration of toluene were kept constant and at the appropriate time the UV light is switched on and reaction is started. Concentration of toluene at the reactor exit is monitored as function of time. Thus, negative step input (exit concentration of toluene decreases in time) simulates non-stationary state of CSTR. Experiments were made with a constant flow of reaction mixture (v_0 = 41 cm^3 min^{-1}) for different length of reactor (1, 3/4, 2/3, 1/2 and 1/3), corresponding to exposed (illuminated) catalyst surface. A series of measurements with different recirculation ratios (60, 49, 37, 24, 12 and 0) are performed. Assuming ideal mixing and first order reaction rate, the final reactor model consists of:

$$r_A = k_A c_A = k_g \cdot a \cdot (c_{A,s} - c_A) \tag{23}$$

$$\frac{\mathrm{d}c_A}{\mathrm{d}t} = \frac{1}{\tau}(c_{A,0} - c_A) - k_g a (c_A - c_{A,s}) \tag{24}$$

Initial conditions are:

$$\begin{aligned} t < 0, \quad & r_A = 0 \\ t \geq 0, \quad & r_A = const. \end{aligned} \tag{25}$$

There are three parameters in the abovementioned model, *kg*, *kA* and , which can be included in parameter estimation. However, in previous experiments with the step input disturbances in the same system without reaction, the residence time, was estimated as parameter that characterize the deviation from ideal mixing. For each experiment with different recirculation ratio this value was applied as constant in the model equations. Moreover, a reaction rate constant in a kinetic model, k_A should not change and it is estimated from the previous experiments in the tubular reactor. Mean value of reaction constant is used as fixed parameter, k_A = 1.02 min^{-1}. Thus, only one parameter, mass transfer coefficient, k_g was estimated applying the model, Eqs. (23)–(25). The higher value of this coefficient is found with the higher recirculation ratio. This is a clear argument that higher velocity of fluid along the surface of catalyst influences global rate of photocatalytic reaction. The results of estimation of parameter k_g, together with previous estimated parameters k_A and τ_m are presented in Table 5. Fig. 7 displays the results of testing the model, Eqs. (23)–(25) for two experiments with different recirculation ratios and with the same area of catalytic layer. The agreement is quite good and as can be seen, the smaller recirculation ratio gives a little steeper curve, which indicates that the dispersion in the reactor slightly decreased.

Table 5 – The results of testing the model of CSTR reactor at different residence times (area of the catalytic layer). Volume rate, v_0 = cm^3 min^{-1}, length of reactor, L = 26 cm.

Recirculation ratio/R	τ_m/min	k_g/cm min^{-1}	SD × 10^3
61	7.35	4.23	4.115
49	5.21	2.52	3.571
37	3.28	1.92	5.413
24	2.26	1.04	1.032
12	1.72	0.87	3.852
k_A = 1.02 min^{-1}			

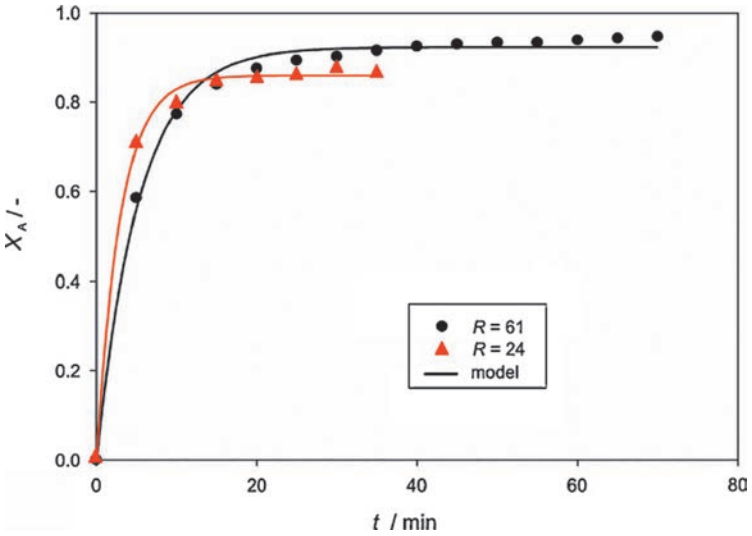


Fig. 7 – Result of testing at different recirculation ratios. Residence time corresponds to the full length of reactor.

4. Conclusion

It is evident from the agreement with the models (Figs. 2 and 3) that the passage of reaction mixture through the reactor is complex and the proposed flow models cannot fully approximate the real flow in tubular reactors (*Model I and II*). Additional understanding of the hydrodynamics is obtained by modelling using computing fluid dynamics (*Model III*). CFD for axialsymmetric flow indicates that design of the reactor with radial introduction of the reaction mixture (U-shape) is better solution than design with the axial introduction (I-shape). Possible explanation is in better mixing and satisfactory using of the working space (Fig. 4). Tubular reactor model which include reaction and assumes ideal flow (*Model IV*) is most appropriate for photochemical reaction analysis because only estimated constants are reaction rate constants, k_A and k_g (Fig. 5). Analysis of the results in the presence of chemical reactions with the axial dispersion model (*Model V*) showed a relatively small value of the estimated parameter *Pe* (3.67–5.63) and indicate the widespread dispersion in the reactor (Fig. 6). Decrease of the reactor length (i.e. reducing the residence time) leads to smaller dispersion (Table 4.). Comparison of the results of a step response in a reactor with recirculation (*Model VI*) indicates that this system can be described with non-ideal flow with large dispersion (Fig. 7). The mean value of the reaction rate constant for toluene oxidation at different hydrodynamic regimes is obtained (k_A = 1.02 min^{-1}). Based on the calculated root mean square deviation, the good agreement of experimental data with the proposed flow model is obtained. Finally, it can be concluded that the reactor hydrodynamics has important influence on the final form of mathematical model useful for detailed description of reactor behaviour.

References

[1] Alberici, R.M., Jardim, W.F., 1997. Photocatalytic destruction of VOCs in the gas-phase using titanium dioxide. Applied Catalysis B: Environmental 14, 55–68.

[2] Bouazza, N., Lillo-Rodenas, M.A., Linares-Solano, A., 2008. Enhancement of the photocatalytic activity of pelletized TiO_2 for the oxidation of propene at low concentration. Applied Catalysis B: Environmental 77, 284–293.

[3] Bouzaza, A., Vallet, C., Laplanche, A., 2006. Photocatalytic degradation of some VOCs in the gas phase using an annular flow reactor. Determination of the contribution of mass transfer and chemical reaction steps in the photodegradation process. Journal of Photochemistry and Photobiology A: Chemistry 177, 212–217.

[4] Dezhi, S., Sheng, C., Chung, J.S., Xiaodong, D., Zhibin, Z., 2005. Photocatalytic degradation of toluene using a novel flow reactor with Fe-doped TiO_2 catalyst on porous nickel sheets. Photochemistry and Photobiology 81, 352–357.

[5] d'Hennezel, O., Pichat, P., Ollis, D.F., 1998. Benzene and toluene gas-phase photocatalytic degradation over H_2O and HCl pretreated TiO_2: by-products and mechanisms. Journal of Photochemistry and Photobiology A: Chemistry 118, 197–204.

[6] Demeestere, K., Visscher, A.D., Dewulf, J., Leeuwen, M.V., Langenhove, H.V., 2004. A new kinetic model for titanium dioxide mediated heterogeneous photocatalytic degradation of trichloroethylene in gas-phase. Applied Catalysis B: Environmental 54, 261–274.

[7] Denny, F., Scott, J., Pareek, V., Peng, G.D., Amal, R., 2009. CFD modelling for a TiO_2-coated glass-bead photo reactor irradiated by optical fibres: Photocatalytic degradation of oxalic acid. Chemical Engineering Science 64, 1695–1706.

[8] Dibble, J.A., Raupp, G.B., 1992. Fluidized-bed photocatalytic oxidation of trichloroethylene in contaminated air streams. Environmental Science and Technology 26, 492–495.

[9] Du, P., Carneiro, J.T., Moulijn, J.A., Mul, G., 2008. A novel photocatalytic monolith reactor for multiphase heterogeneous photocatalysis. Applied Catalysis A-General 334, 119–128.

[10] Ertl, G., Knözinger, H., Weitkamp, J., 1997. Handbook of Heterogeneous Catalysis. Wiley-VCH, Weinheim.

[11] Fogler, H.S., 1986. Elements of Chemical Reaction Engineering. Prentice-Hall, New Jersey.

[12] Ge, H., Chen, G., Yuan, Q., Li, H., 2005. Gas phase catalytic partial oxidation of toluene in a microchannel reactor. Catalysis Today 110, 171–178.

[13] Huang, H., Li, W., 2011. Destruction of toluene by ozone-enhanced photocatalysis: performance and mechanism. Applied Catalysis B: Environmental 102, 449–453.

[14] Ibhadon, A.O., Arabatzis, I.M., Falaras, P., Tsoukleris, D., 2007. The design and photoreaction kinetic modeling of a gas-phase titania foam packed bed reactor. Chemical Engineering Journal 133, 317–323.

[15] Ichiura, H., Kitaoka, T., Tanaka, H., 2003. Removal of indoor pollutants under UV irradiation by a composite TiO_2-zeolite sheet prepared using a papermaking technique. Chemosphere 50, 79–83.

[16] Iguchi, Y., Ichiura, H., Kitaoka, T., Tanaka, H., 2003. Preparation and characteristics of high performance paper containing titanium dioxide photocatalyst supported on inorganic fiber matrix. Chemosphere 53, 1193–1199.

[17] Imoberdorf, G.E., Irazoqui, H.A., Alfano, O.M., Cassano, A.E., 2007. Scaling-up from first principles of a photocatalytic reactor for air pollution remediation. Chemical Engineering Science 62, 793–804.

[18] Jeong, J., Sekiguchi, K., Sakamoto, K., 2004. Photochemical and photocatalytic degradation of gaseous toluene using short-wavelength UV irradiation with TiO_2 catalyst: comparison of three UV sources. Chemosphere 57, 663–671.

[19] Keller, V., Bernhardt, P., Garin, F., 2003. Photocatalytic oxidation of butyl acetate in vapor phase on TiO_2, Pt/TiO_2 and WO_3/TiO_2 catalysts. Journal of Catalysis 215, 129–138.

[20] Kim, S.B., Hwang, H.T., Hong, S.C., 2002. Photocatalytic degradation of volatile organic compounds at the gas–solid interface of a TiO_2 photocatalyst. Chemosphere 48, 437–444.

[21] Ku, Y., Ma, C.-M., Shen, Y.-S., 2001. Decomposition of gaseous trichloroethylene in a photoreactor with TiO_2-coated nonwoven fiber textile. Applied Catalysis B: Environmental 34, 181–190.

[22] Lasa, H.D., Serrano, B., Salaices, M., 2005. Photocatalytic Reaction Engineering. Springer, London.

[23] Lasa, H.I.d., Rosales, B.S., 2009. Photocatalytic Treatment of Air: From Basic Aspects to Reactors.

[24] Levenspiel, O., 1999. Chemical Reaction Engineering. Wiley.

[25] Lim, T.H., Kim, S.D., 2005. Photocatalytic degradation of trichloroethylene (TCE) over TiO_2/silica gel in a circulating fluidized bed (CFB) photoreactor. Chemical Engineering and Processing 44, 327–334.

[26] Maira, A.J., Coronado, J.M., Augugliaro, V., Yeung, K.L., Conesa, J.C., Soria, J., 2001. Fourier transform infrared study of the performance of nanostructured TiO_2 particles for the photocatalytic oxidation of gaseous toluene. Journal of Catalysis 202, 413–420.

[27] Mendez-Roman, R., Cardona-Martinez, N., 1998. Relationship between the formation of surface species and catalyst deactivation during the gas-phase photocatalytic oxidation of toluene. Catalysis Today 40, 353–365.

[28] Mohseni, M., Taghipour, F., 2004. Experimental and CFD analysis of photocatalytic gas phase vinyl chloride (VC) oxidation. Chemical Engineering Science 59, 1601–1609.

[29] Nauman, E.B., 2008. Chemical Reactor Design, Optimization and Scaleup. John Willey& Sons.

[30] Nevers, N.D., 1995. Air Pollution Control Engineering. McGraw Hill, Singapore.

[31] Nicolella, C., Rovatti, M., 1998. Mathematical modeling of monolith reactors for photocatalytic oxidation of air contaminants. Chemical Engineering Journal 69, 119–126.

[32] Peill, N.J., Bourne, L., Hoffmann, M.R., 1997. Iron(III)-doped Q-sized TiO_2 coatings in a fiber-optic cable photochemical reactor. Journal of Photochemistry and Photobiology A: Chemistry 108, 221–228.

[33] Prieto, O., Fermoso, J., Irusta, R., 2007. Photocatalytic degradation of toluene in air using a fluidized bed photoreactor. International Journal of Photoenergy 2007, 8.

[34] Queffeulou, A., Geron, L., Schaer, E., 2010. Chemical Engineering Science 65, 5067–5074.

[35] Quici, N., Vera, M.L., Choi, H., Li Puma, G., Dionysiou, D.D., Litter, M., Destaillats, H., 2010. Applied Catalysis B: Environmental 95, 312–319.

[36] Raupp, G.B., Alexiadis, A., Hossain, M.M., Changrani, R., 2001. First-principles modeling, scaling laws and design of structured photocatalytic oxidation reactors for air purification. Catalysis Today 69, 41–49.

[37] Romero-Vargas Castrillón, S., Ibrahim, H., Lasa, H.D., 2006. Flowfield investigation in a photocatalytic reactor for air treatment (Photo-CREC–air). Chemical Engineering Science 61, 3343–3361.

[38] Salvadó-Estivill, I., Hargreaves, D.M., Li Puma, G., 2007a. Evaluation of the intrinsic photocatalytic oxidation kinetics of indoor air pollutans. Environmental Science and Technology 41, 2028–2035.

[39] Salvadó-Estivill, I., Brucato, A., Li Puma, G., 2007b. Two-dimensional modeling of a flat-plate photocatalytic reactor for oxidation of indoor air pollutants. Industrial and Engineering Chemistry Research 46, 7489–7496.

[40] Schiesser, W.E., 1991. The Numerical Method of Lines: Integration of Partial Differential Equations. Academic Press, San Diego.

[41] Sekiguchi, K., Yamamoto, K., Sakamoto, K., 2008. Photocatalytic degradation of gaseous toluene in an ultrasonic mist containing TiO_2 particles. Catalysis Communications 9, 281–285.

[42] Taghipour, F., Mohseni, M., 2005. CFD simulation of UV photocatalytic reactors for air treatment. AIChE Journal 51, 3039–3047.

[43] Tomasic, V., Jovic, F., Gomzi, Z., 2008. Photocatalytic oxidation of toluene in the gas phase: modelling an annular photocatalytic reactor. Catalysis Today 137, 350–356.

[44] Zhang, Y., Yang, R., Zhao, R., 2003. A model for analyzing the performance of photocatalytic air cleaner in removing volatile organic compounds. Atmospheric Environment 37, 3395–3399.

[45] Zuo, G.-M., Cheng, Z.-X., Chen, H., Li, G.-W., Miao, T., 2006. Study on photocatalytic degradation of several volatile organic compounds. Journal of Hazardous Materials B128, 158–163.

doi: 10.15255/CABEQ.2014.19399

Chemometric Versus Random Forest Predictors of Ionic Liquid Toxicity*

Želimir Kurtanjek**

Faculty of Food Technology and Biotechnology, University of Zagreb, Croatia

The objective of this work is comparative analysis of standard chemometric and decision tree(s) models for prediction of biological impact of ionic liquids (ILs) for various combinations of cations and anions. The models are based on molecular descriptors for combinations of the following cations: imidazole, pyridinium, quinolinium, ammonium, phosphonium; and anions: BF_4, Cl, PF_6, Br, CFNOS, NCN_2, $C_6F_{18}PBF_4$, Cl, PF_6, Br, CFNOS, $C_6F_{18}P$. Derived data matrix is decomposed by singular value decomposition of the cation and anion matrices into corresponding first ten components each accounting for 99,5 % of the corresponding total variances. Biological impact data, i.e. molecular level toxicity, are based on acetylcholinestarase inhibition experimental data provided in MERCK Ionic Liquids Biological Effects Database. Applied are the following models: principal component regression (PCR), partial least squares (PLS), and decision tree(s) model. The model performances are compared by ten fold validation. Obtained are the following Pearson regression coefficients R^2: PCR 0.62, PLS 0.64, and for decision tree forest RFDT 0.992. The decision tree(s) models significantly outperformed chemometric models for numerical predictions of E_{50} concentrations and the classification of ILs into four level of toxicities.

Key words:
ionic liquids, toxicity, chemometrics, decision tree

*The article was originally published in the journal „Chemical and Biochemical Engineering Quarterly Journal“ 28 (4) (2014) 459-463 by the publisher Croatian Association of Chemical Engineers and Technologists . The article has been approved for scholarly and non-commercial use.
**retired professor of chemical engineering, zelimir.kurtanjek@gmail.com

1. Introduction

Interest in ionic liquids ILs stems from their unique solvent properties and potential process "self-containment". Its applications in chemical processes and biotransformations provide possibility for clean manufacturing ("green technology"). Besides their solvent and extraction functions ILs also exhibit synergy effects with catalysts (enzymes) yielding higher production productivity. Theoretically there is a limitless number of possible ILs with a very broad range of physical and chemical properties. Research on ILs has become one of the most interesting application research areas in novel catalytic synthesis, biofuel production from agricultural wastes, integration of chemical and enzyme reactors with separation processes, polymerization, nanotechnology, enzyme-catalysis, composite preparation and renewable resource utilization[1-3]. Especially is interesting use of microreactors for ionic liquid synthesis and possibly as production systems for integrated biotransformations and product separation[4]. However, recent question of ILs eco-toxicity and their degradability have also been raised.

Analysis of their versatile structure is formally viewed as a combinatorial problem which can be effectively accounted by computers. The object of this work is to apply computer modeling by chemometric methodology and decision tree algorithm for predicting continuous variables, such as toxicity level concentration E_{50} and level classification, based on choice of cation and anions structure and their chemical compositions. Predictions are ILs physical properties are based on literature published data an internet available NIST and MERCK databases of physical properties and cytotoxicity[5-7].

The main objective of this work is in inferring of rules and patterns implicitly contained in a set of chemical structures and molecular descriptors. Applied is a supervised learning algorithm with target sets for continuous and classification properties reveal relationships between molecular descriptors.

2. Experimental

The chemical formula of each ion is recorded in SMILES and MOL format and evaluated for corresponding molecular descriptors[8]. Hence, each combination of ions for a specific IL is represented by 2x797 data points. Since numerical values of molecular descriptors cover a range of numerical orders of magnitude, each descriptor is autoscaled based on the sample average and the corresponding standard deviation. For the selected cations the transformation is:

$$\mathbf{X}_C \leftarrow \frac{\mathbf{X}_C - \bar{\mathbf{X}}_C}{\sigma(\mathbf{X}_C)} \tag{1}$$

Similarly, molecular descriptors for the selected anions is transformed accordingly:

$$\mathbf{X}_A \leftarrow \frac{\mathbf{X}_A - \bar{\mathbf{X}}_A}{\sigma(\mathbf{X}_A)} \tag{2}$$

Obtained matrices of the autoscaled descriptors are analyzed for their mutual interrela-tionships. For cations and anions data matrices average Pearson correlation of R^2 =0.4 is obtained, which is significant considering the large number of samples (ions). Due to high co-linearity between various molecular descriptors the data matrices are decomposed into series of partial components by application of singular value decomposition of the corresponding anion $\mathbf{X}_A$ and cation $\mathbf{X}_C$ covariances by solving the eigenvalue problems:

$$\left(\mathbf{X}_C^T \mathbf{X}_C\right)\mathbf{v}_{C,i} = \lambda_{C,i}\,\mathbf{v}_{C,i} \quad j = 1,2\cdots M \tag{3}$$

$$\left(\mathbf{X}_A^T \mathbf{X}_A\right)\mathbf{v}_{A,i} = \lambda_{A,i}\,\mathbf{v}_{A,i} \quad j = 1,2\cdots M \tag{4}$$

Decomposed matrices, $\mathbf{P}_A$ and $\mathbf{P}_C$, are defined by the corresponding eigenvectors $\mathbf{v}_A$ and $\mathbf{v}_C$, and the contributions of individual partial decompositions are evaluated by the ratios of squares corresponding eigenvalues λ_A and λ_A to the number of descriptors *M*.

$$\mathbf{P}_C = \mathbf{X}_C \cdot \left(\mathbf{v}_{C,1} \middle| \mathbf{v}_{C,2} \middle| \cdots \mathbf{v}_{C,K}\right) \tag{5}$$

$$\mathbf{P}_A = \mathbf{X}_A \cdot \left(\mathbf{v}_{A,1} \middle| \mathbf{v}_{A,2} \middle| \cdots \mathbf{v}_{A,K}\right) \tag{6}$$

Based on preselected level of 99.5 % of the total variance the first ten eigenvectors for each data set are chosen.

3. Results and Discussion

Compared are the chemometric and decision tree models for regression and prediction of concentration E_{50} and toxicity classification for inhibition of acetylchlorinestarase inhi-bition experimental data provided in MERCK Ionic Liquids Biological Effects Database[7]. The model input data are the target values of molecular descriptor projections. The chemometric models are linear models, and here are applied

based on their expected robustness and improved prediction when compared to classical least squares multivariate models[9-12]. The first tested model is Principal Component Regression (PCR) given by Eq. (7).

$$\mathbf{Y} = \text{²}_A \cdot \mathbf{P}_A + \text{²}_C \cdot \mathbf{P}_C + \mathbf{E} \tag{7}$$

The statistical evaluation and analysis of the model parameters are performed by the algorithms provided by R open source software[16] and STATISTICA[17]. Applied is ten-fold cross validation within the training set of samples and also a validation with the data set that has not been used during the modeling phase. The model "quality" for prediction of E_{50} concentration is relatively "poor" with $R^2 = 0.62$ presented in Fig. 1.

The second tested model is Partial Least Squares (PLS) which is to improve the predictions by separate decomposition of the input and output data sets (Eq 8-9).

$$\mathbf{X} = \mathbf{T} \cdot \left(\mathbf{P}_C \middle| \mathbf{P}_A \right)^T + \mathbf{E}_X \tag{8}$$

$$\mathbf{Y} = \mathbf{U} \cdot (\mathbf{Q})^T + \mathbf{E}_Y \tag{9}$$

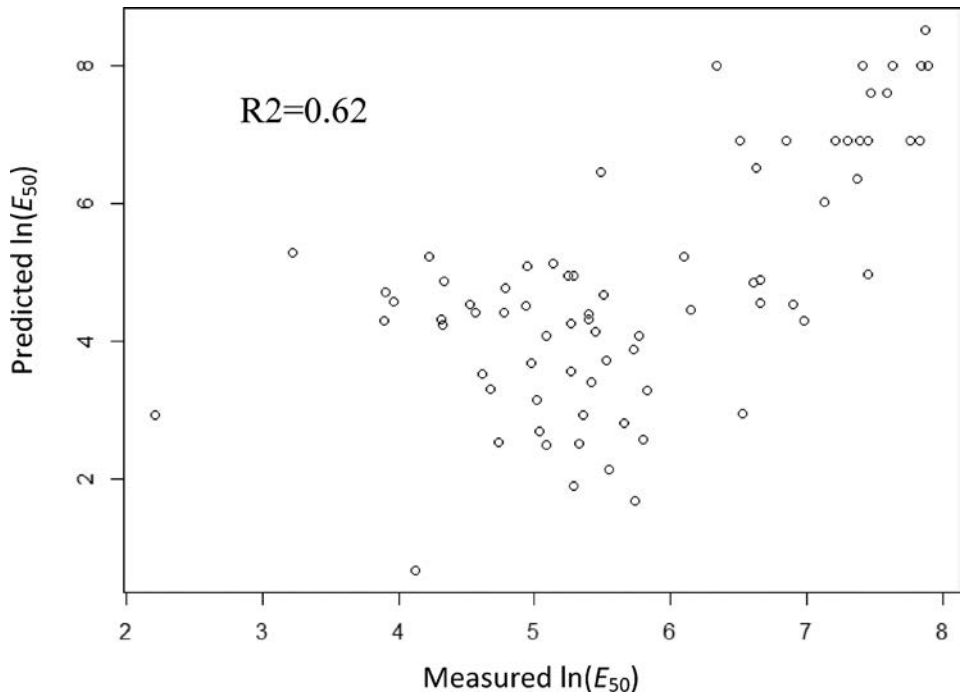


Fig. 1 – Comparison between the test samples for measured ln(E_{50}) concentrations and the principal component regression model predictions

The predictive model is built by regression between the inner projections **T** and **U**:

$$\mathbf{U} = \mathbf{T} \cdot \beta + \mathbf{E} \tag{10}$$

The PLS model predictions on the test data slightly improved yielding $R^2 = 0.64$ and are presented in Fig. 2.

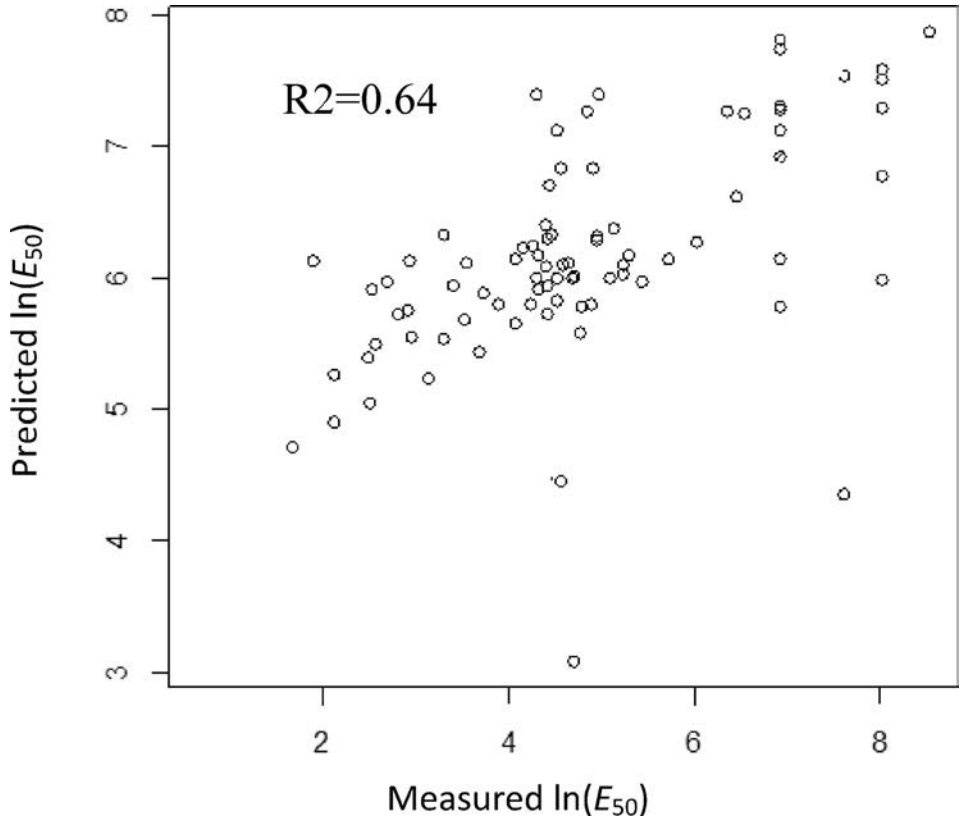


Fig. 2 – Comparison between the test samples for measured ln(E_{50}) concentrations and the partial least squares model predictions

The obtained relatively poor predictions of E_{50} by the chemometric models is in contrast to good predictions for some of ILs physical properties, for example as for viscosity, given in literature[12-14]. A possible explanation is due to high dispersion of the experimental data E_{50} involved in measurement of biological effect of ILs.

In order to elevate the modelling assumption on continuity and linearity between molecular descriptors and biological effects applied are decision tree (DT) and ran-

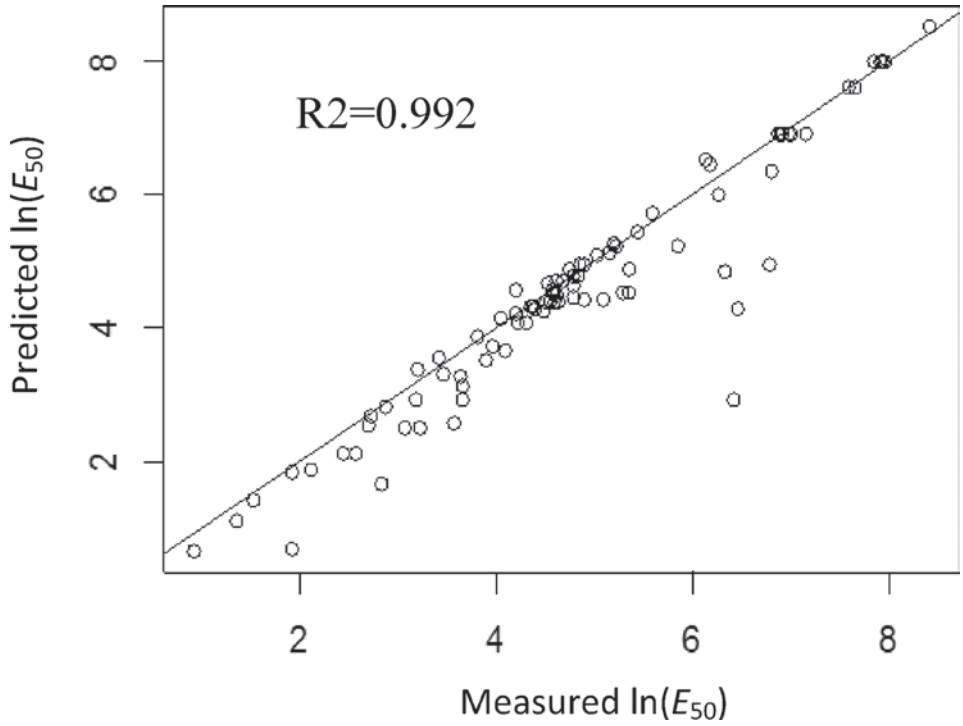


Fig. 3 – Comparison between the test samples for measured ln(E_{50}) concentrations and the random forest model predictions.

dom forest (RF) models[11,15,18]. These are nonparametric models and are not based on assumed functional relationships between the input and output data. Main objective of decision tree model is a supervised procedure of step-wise classification of input data by binary split into subsets for "improved" or more significant information content (information gain). It is obtained by minimisation of Gini index or pattern entropy. Produced models are not given in a closed mathematical form, but as a set of logical statements which can be easily represented in a graphical form as a tree of stepwise decisions. When a DT model is used for regression then the numerical range of output data is approximated by pseudo classes for assumed precision of regression predictions. Here is applied Breiman and Cutler[15] and the tree GA evolution algorithm available in R software system and tree plotting[16-18].

$$\mathbf{Y}_{DT}^{P} = DT\left(\mathbf{Y}, \mathbf{P}_C, \mathbf{P}_A\right) \tag{11}$$

Single decision tree prediction models tend to be biased but modelling can be improved by re-initialization of collections of trees by randomisation of the split algorithm and producing random forest. Prediction of a random forest is obtained by aggregation of individual trees with weighted response corresponding to individual tree cross-validation.

$$\mathbf{Y}_{RF}^{P} = RF\left(\mathbf{Y}, \mathbf{P}_C, \mathbf{P}_A\right) \tag{12}$$

Modelling results are presented in Fig. 3-4. Prediction of $\ln(E_{50})$ by the random forest model is greatly improved with Pearson correlation $R^2 = 0.992$. Individual decision tree for classification of ILs toxicity is depicted in Fig. 4. Here are for acetycholinesterase adopted the following classes: low (L, $E_{50} < 10$ μmol·L^{-1}), medium (M), $E_{50} \in [10 - 100$ μmol L$^{-1}]$, high (H) $E_{50} \in [100 - 1000$ μmol L$^{-1}]$, and very high (VH) $E_{50} > 1000$ μmol L^{-1}according to MERCK classification[7]. The advantage of application of uncorrelated principal components of the molecular descriptor sets has resulted in a simple and transparent model.

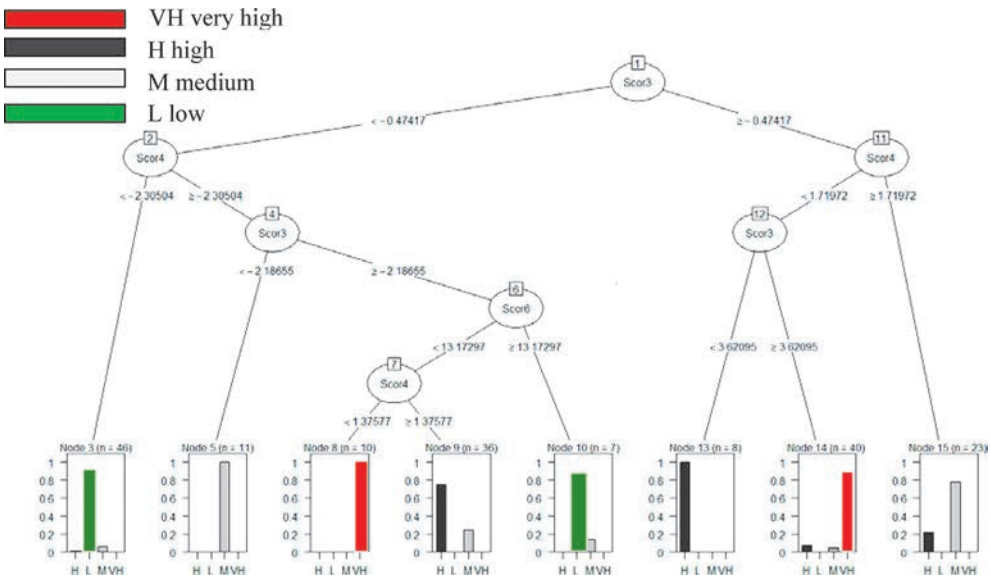


Fig. 4 – Decision tree model for predictions based on classification of $\ln(E_{50})$ into VH (very high), H (high), M (medium) and L (low) ILs toxicity categories.

4. Conclusions

Applied are chemometric and decision tree models of ILs toxicity based on their molecular descriptors. Toxicity criteria is based on E_{50} concentrations for inhibition of acetylcholinesterase, In view of very large of molecular descriptors their colinearity was investigated and was found significant average correlation $R^2 \oplus 0.4$. In order to simplify and obtain robust models the matrices of cation and anion descriptors are projected to the corresponding spaces of the first ten eigenvectors resulting into about 99,95 of variance (data dispersion content).

Application of chemometric models, partial component regression and partial least squares, resulted in limited quality of prediction on test sets with regression coefficients R^2 of 0.62 and 0.64. However, application of decision tree and random forest models significantly improved quality of prediction with $R^2 = 0.992$. Due to orthogonalisation of training patterns derived are simple and transparent decision trees.

Practical application of the derived models is in potential use as a part of a feedback loop for inverse design of new ILs for specific (tailored) new technology process needs.

Acknowledgements

This work was supported by Ministry of Science, Education and Sports of Republic of Croatia, project 058-1252086-0589

References

[1] Kokorin A., , Ionic Liquids: Theory, Properties, New Approaches, InTech, Rijeka, Croatia 2011

[2] Plechkova N.V., Seddon K.R., Izgorodina E.I., Theoretical Approaches to Ionic Liquids: From Past History to Future Directions, Ionic Liquids Uncoiled: Critical Expert Overviews, Editors: Plechkova N.V., Seddon K.R., Wiley Online Library (2013), DOI: 10.1002/9781118434987

[3] SDEWES, 2012, Fraunhofer-Allianz Annual Report, <www.allianz.com/en/investor_relations/results_reports/annual-reports.html>, accessed on 14/12/2013

[4] Cvjetko M., Synthesis, application in biotransformations and cytotoxycity of selected imidazolium-based ionic liquids, Doctoral Thesis, (in Croatian) University of Zagreb, Faculty of Food Technology and Biotechnology, Zagreb, Croatia, (2012)

[5] Suojiang Z., Lu X., Zhou Q., Li X., Zhang X., Li S., "Ionic Liquids, Physichochemical Properties", Elsevier, Amsterdam, The Netherlands, 2009

[6] NIST, Ionic Liquids Database Standard Reference Database #147, <ilthermo.bouder.-nist.gov/ILThermo/mainmenu.uix>, accessed on 18/07/2013

[7] The UFT/ Merck Ionic Liquids Biological Effects Database, <www.il-eco.uft.uni-bremen.de>, accessed on14/12/2013.

[8] Chun Wei Yap, *J. Comp. Chem*., **32**(2010) 1466-1474

[9] Brereton G.R., *J. Chemometrics*, **28** (2014) 749-760

[10] Varmuza K., Filzmoser P., "*Multivariate Statistical Analysis in Chemometrics",* CRC Press, Baton Rouge, Louisiana, USA, 2009.

[11] Wehrens R., *Chemometrics with R*, 2011, Springer, New York, USA, 2011

[12] Fatemi M.H., Izadiyan P.,*Chemosphere*, **84**, (2011) 553-563.

[13] Svetnik V., Liaw A., Tong C., Culberson J.C., Sheridan R.P., Feuston B.P., 2003, *J. Chem. Inf. Comput. Sci.,* **43** (2003) 1947–1958

[14] Kurtanjek, Ž., Proceedings of "*24-th European Symposium on Computer Aided Process Engineering*", Ed. Klemeš J.J., Varbanov P.S., Liew P.Y, Budapest, Hungary, 15-19 June (2014) 127

[15] Breiman L., Cutler A.,< www.stat.berkeley.edu/~breiman/ RandomForests>, accessed on 14/12/2013

[16] T. Grubinger, Zeileis, K.P. Pfeiffer, Evolutionary Learning of Globally Optimal cation and Regression Trees, Journal of Statisical Software, 61(1), 1-29 (2014)

[17] R Development Core Team, R: A language and environment for statistical computing. R, Vienna, Austria,<www.R-project.org>, accessed on 14/12/2013

[18] StatSoft, Inc. STATISTICA, v.10.< www.statsoft.com>, accessed on 14/12/2013.

[19] Therneau T., Atkinson B., Ripley B., 2013, CRAN – Packagge rpart, <cran.rproject.org/web/packages/rpart/index.html>, accessed on 14/12/2013.

doi: n/a

Determination of the Elastic Constants of a Plain Woven Fabrics by Tensile Test in Various Directions*

Željko Penava, Diana Šimić Penava, Željko Knezić**

Croatian Academy of Engineering

In this paper the values of elastic constants of woven fabrics for different angles of extension direction were analyzed. Four types of fabric samples of different raw material composition and the same type of weave were tested under tensile forces in seven directions oriented with 15° increment with respect to the weft direction. Elasticity modulus and Poisson's ratio of woven fabrics were determined experimentally in the laboratory. Based on the experimentally obtained values, theoretically calculated elastic constants for arbitrarily chosen fabric directions were calculated. A good agreement between experimental results and the calculated obtained values of the elastic constants was shown, so the theoretical equations can be used with high accuracy to calculate the elastic constants of the fabric in various directions. Therefore, the measurements need to be implemented when the tensile force acting on the fabric only in the warp, weft and at an angle of 45°.

Key words:

woven fabric, anisotropic, modulus of elasticity, Poisson's ratio, elastic constants

1. Introduction

Anisotropy is a characteristic of most materials, especially woven fabrics. The impact of the direction of action of external load (tensile force) on the properties of

* The article was originally published in the journal "Fibres & Textiles in Eastern Europe" 2014, Vol. 22, No. 2 (104): 57-63 by the publisher Institute of Biopolymers and Chemical Fibres, Łódź, Poland. The article has been approved for scholarly and non-commercial use.

** zpenava@ttf.hr

the fabric is enormous, and is frequently examined (Lekhnitskii, 1981). In a biaxial woven structure two main directions are defined: longitudinal (warp) and transverse (weft). When load acts in one of the main directions, the extension which causes a rupture is minimal, and the associated modules assume the maximum value (Sengupta et al., 1972). When the angle of an external load (tensile force) changes, elastic constants change too.

Although measurements of strain and stress at fabric rupture during force action in the longitudinal or transverse direction are dealt with in most experimental studies, the behavior of fabric and deformation are equally important when changing the inclination angle of tensile force (load), especially in technical textiles (Shanahan et al., 1978., Kovar et al., 2009). For example, in parachute materials as well as in composites, stress and strain can occur in different directions of force action. When the action angle of external load changes, the elastic constants of the fabric also change. The amount of breaking force and elongation of the fabric are determined experimentally in laboratory tests of the fabric. Using these experimentally obtained values, modulus of elasticity is determined. Therefore, it is necessary to better understand the mechanical properties of fabric behavior. The interaction between the yarns in woven fabrics under tensile force play a significant role. Although the application of textiles in different industries is growing, especially in the development of composite materials, understanding the behavior of the mechanical properties of the fabric is still limited. Kilby was among the first to start studying the mechanical properties of the material under the action of tensile load, 1937 (Pierce, 1937). He begins with the classical theory of elasticity with the assumption that the fabric is an anisotropic material with two planes of symmetry (Kilby, 1963). He defined Poisson's ratio and gave the dependence of the modulus of elasticity of fabric in relation to the direction of action of tensile forces. The tensile properties of fabrics in an arbitrary direction of tensile force were also measured. The method for measuring anisotropic tensile properties of fabrics is called "uniaxial test method" or the method of force action in only one direction (Bao et al., 1997., Bao et al., 1997., Bassett et al., 1999., Loyd, Hearle, 1997).

The aim of this work was to analyse the influence of anisotropy on the elastic constants of a plain woven fabrics in various directions and to determine the degree of agreement between experimental results and calculated obtained values.

2. Elastic constants of anisotropic materials

The fabric is an elastic anisotropic material. Fabrics as anisotropic materials are a special type of porous materials and can be treated as non-uniform mixtures of

fibers and air (Kuwazuru, Yoshikawa, 2004). The impact of external forces on the fabric causes them to change their shape and volume. Theoretical analysis of fabric behavior is often very complex, so that the experimental verification of theoretical predictions for them is more important than for other materials (Kuwazuru, Yoshikawa, 2004).

2.1 Woven fabrics as anisotropic materials

Woven fabrics and textile materials are generally inhomogeneous, anisotropic and discontinuous objects so that even under moderate loads they have large deformations and displacements in propagation planes (Hu, 2004). Determination of mechanical properties and prediction of fabric behavior during the production process and finally in use is a very important part of textile science.

These issues have been dealt with over the years by many researchers, and many complex measurement systems for measuring various mechanical characteristics of fabric have been designed (Chen et al., 2007., Kovar, 2003). As these fabric features are mainly explored regarding elasticity, specifically under low loads, it is very difficult to achieve sufficient accuracy in determining the degree of deformation and spatial displacement of material (Postle et al., 1988). It is necessary to take care that the measuring system has no impact on the test specimen; for this purpose optical measurement methods or processing video recordings are used nowadays Zouari et al., 2010). Such a procedure was applied to the experimental part of this work.

2.2 Elastic constants of anisotropic material in the direction of the major axes

Special type of anisotropic materials having two mutually perpendicular planes of elastic symmetry are called orthotropic materials. These planes of elastic symmetry are planes of orthotropy, and their cross sections are axes of orthotropy. The most widely used orthotropic materials are: wood, plywood, glass reinforced plastics, laminates and other composite materials. These materials are mainly used in the form of plates with plane state of stress (Herman, 2008., Pan, Yoon, 1996). The element of an orthotropic plate (woven fabric) with plane state of stress is shown in Figure 1.

The x- and y-axes coincide with the axes of orthotropy *T*, *L*, and the *z*-axis is perpendicular to the plane of the element. To get the relationship between stress and strain, we must determine modulus of elasticity (Young's modulus) E_i and Poisson's ratio ν_{ij} which are determined experimentally in laboratory tests. Constants E_i and ν_{ij} are also called elastic constants (constants of elasticity), while the technical

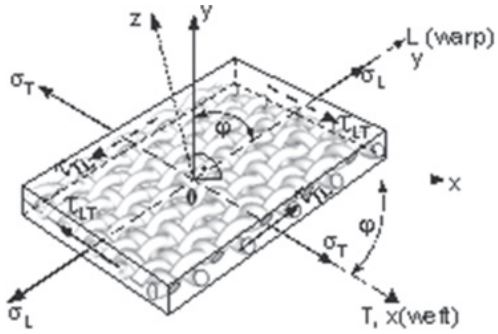


Fig. 1 – Element of an orthotropic plate – woven fabric

literature uses the term engineering elastic constants. For the plane state of stress of an elastic, homogeneous, isotropic material in a linear region (linear relationship between stress and strain) Hooke's law is applied, which is:

$$\begin{Bmatrix} \varepsilon_T \\ \varepsilon_L \\ 2\varepsilon_{TL} \end{Bmatrix} = \begin{bmatrix} \frac{1}{E_T} & -\frac{\nu_{LT}}{E_T} & 0 \\ -\frac{\nu_{TL}}{E_L} & \frac{1}{E_L} & 0 \\ 0 & 0 & \frac{1}{G_{LT}} \end{bmatrix} \cdot \begin{Bmatrix} \sigma_T \\ \sigma_L \\ \tau_{TL} \end{Bmatrix} \tag{1}$$

Equation (1) represents the matrix form of Hooke's law. In these expressions the index T refers to the transverse or transverse axis (weft), with $\varphi = 0°$, and the index L in the longitudinal or longitudinal axis (warp), at $\varphi = 90°$, Figure1. Transverse and longitudinal elasticity modulus are denoted by E_T and E_L and shear modulus is denoted by G_{TL} and G_{LT}. Poisson' ratios are denoted by ν_{LT} and ν_{TL}, respectively. Furthermore, in the expression (1) σ_T represents normal stress on the plane which is perpendicular to the direction of the transverse axis T, σ_L is the normal stress on the plane which is perpendicular to the direction of the longitudinal axis L and $\tau_{i,j}$ is shear (tangential) stress. The first index *i* indicates the direction of the vertical line to the plane in which stress acts, and the second index *j* indicates the direction of stress and the relation $\tau_{ij} = \tau_{TL} = \tau_{LT}$ applies. Detailed description of normal σ_i and shear $\tau_{i,j}$ stresses is given in (Penava, Šimić, 2012). The denotation ε_T represents a relative length deformation in the direction of the T-axis, and ε_L is a relative length deformation in the direction of the L-axis. In the case of an uniaxial state of stress Hooke's law can be written as follows:

$$E = \frac{\sigma}{\varepsilon} = \frac{F}{\varepsilon \cdot b \cdot d}[MPa] \tag{2}$$

where: F – tensile force [N], ε – elongation [%], b – width [mm], d – fabric thickness [mm].

2.3 Poisson's ratio of a woven fabric

When a tensile force acts in one direction on the fabric, in this direction the fabric stretches, and in a direction perpendicular to the direction of force, the fabric contracts. This phenomenon describes the Poisson's ratio. Poisson's ratio is one of the fundamental properties of any structural material including fabric. This coefficient determines important mechanical characteristics of fabrics in many applications, including a variety of composite systems containing textiles as a structural element. Values of Poisson's ratio for the fabrics are different from those of standard engineering materials and they can reach values outside the range 0.1-2 (Sun, 2005). Knowing the value of the coefficient is also necessary when creating computer simulations of fabrics and garments. To determine the Poisson's ratio of fabrics, devices for measuring tensile strength are used, and the coefficient is determined in the linear part of the diagram of Hooke's law. The researchers determined the Poisson's ratio in the warp and weft direction of the fabric based on the geometric models of fabric and excluded the impact of the Poisson's ratio of the yarn. In this way, they came to the conclusion that the Poisson's ratio of fabrics results from the interaction between the warp and weft, and can be expressed in terms of structural and mechanical parameters of the system (Jinyun et al., 2010). The physical meaning of Poisson's ratio is shown by expression (3). Longitudinal and transverse strains have an opposite sign.

$$\nu = -\frac{\varepsilon_p}{\varepsilon} \tag{3}$$

Poisson's ratio is defined as a negative value of the ratio of transverse ε_p and longitudinal extension strain ε in relation to the arbitrary action direction of tensile force on the fabric sample.

2.4 Elastic constants of anisotropic material for arbitrarily oriented axes

For an orthotropic and elastic material when the x- and y-axes do not coincide with the axes of orthotropy the anisotropic behaviours under tensile loads can be written in matrix form by Hooke's law (4). Elastic constants E_x, E_y, G_{xy}, ν_{xy}, α_x, α_y in the coordinate system x, y are estimated by equations (5-10) for the transformation of the elastic constants as in (Alfirević, 2003).

$$\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ 2\varepsilon_{xy} \end{Bmatrix} = \begin{bmatrix} \frac{1}{E_x} & -\frac{\nu_{yx}}{E_x} & \alpha_x \\ -\frac{\nu_{xy}}{E_y} & \frac{1}{E_y} & \alpha_y \\ \alpha_x & \alpha_y & \frac{1}{G_{xy}} \end{bmatrix} \cdot \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} \tag{4}$$

$$\frac{1}{E_x}=\frac{\cos^4\phi}{E_T}+\frac{\sin^4\phi}{E_L}+\left(\frac{1}{G_{TL}}-\frac{2\cdot\nu_{LT}}{E_T}\right)\cdot\cos^2\phi\cdot\sin^2\phi\,, \tag{5}$$

$$\frac{1}{E_y}=\frac{\sin^4\phi}{E_T}+\frac{\cos^4\phi}{E_L}+\left(\frac{1}{G_{TL}}-\frac{2\cdot\nu_{LT}}{E_T}\right)\cdot\cos^2\phi\cdot\sin^2\phi\,, \tag{6}$$

$$\frac{1}{G_{xy}}=\frac{1}{G_{TL}}+\left(\frac{1}{E_T}+\frac{1}{E_L}+\frac{2\cdot\nu_{LT}}{E_T}-\frac{1}{G_{TL}}\right)\cdot 4\cos^2\phi\cdot\sin^2\phi=\frac{1}{G_{yx}}\,, \tag{7}$$

$$\nu_{xy}=\frac{E_x}{E_T}\left[\nu_{LT}-\left(1+2\nu_{LT}+\frac{E_T}{E_L}-\frac{E_T}{G_{TL}}\right)\cdot\cos^2\phi\cdot\sin^2\phi\right], \tag{8}$$

$$\alpha_x=2\left[\left(\frac{1+2\nu_{LT}}{E_T}+\frac{1}{E_L}-\frac{1}{G_{TL}}\right)\cdot\sin^2\phi+\frac{1}{2G_{TL}}-\frac{\nu_{LT}}{E_T}-\frac{1}{E_L}\right]\cdot\cos\phi\cdot\sin\phi\,, \tag{9}$$

$$\alpha_y=2\left[\left(\frac{1+2\nu_{LT}}{E_T}+\frac{1}{E_L}-\frac{1}{G_{TL}}\right)\cdot\cos^2\phi+\frac{1}{2G_{TL}}-\frac{\nu_{LT}}{E_T}-\frac{1}{E_L}\right]\cdot\cos\phi\cdot\sin\phi\,. \tag{10}$$

where: $E_{45}=E_x$, a $\nu_{45}=\nu_{xy}$, when the x-axis closes an angle of 45° with the T-axis. The shear modulus (Alfirević, 2003) is shown in equation (11):

$$G_{TL}=\frac{E_{45}}{2\cdot(1+\nu_{45})}\,. \tag{11}$$

3. Experimental determination of elastic constants of woven fabrics

In the experimental part of the work the magnitudes of tensile forces were determined which are used to calculate modulus of elasticity, elastic coefficients and the Poisson's ratio of the fabric depending on the direction of tensile force to the fabric. For this purpose, classical methods and instruments for testing tensile properties of fabrics were used. The conducted experiment was to confirm that the calculation of the elastic constants for anisotropic material is entirely true for woven fabrics or textile surface materials. The experiment was carried out by measuring fabric spatial deformation under the action of tensile force till rupture, specifically in: warp direction, in weft direction, and under angles of 15°, 30°, 45°, 60°, 75° to the weft.

3.1 Samples and testing apparatus

To carry out this study, four different fabrics of different raw material composition (cotton, wool, wool + lycra, PES) and of the same weave (plain weave) were available. Raw material and structural properties of the tested fabrics are shown in Table 1.

Table 1 – Structural characteristics of the tested samples

Material	100 % cotton	100 % wool	95% wool 5% lycra	100 % polyester (PES)
Warp linear density (tex)	32	50.6	28	32
Weft linear density (tex)	30	47	30	22
Warp density (cm^{-1})	22	26	32	31
Weft density (cm^{-1})	22	18	29	26
Fabric weight (g/m^2)	150.34	234.75	178.22	164.57
Fabric thickness (mm)	0.318	0.568	0.328	0.252

Yarn count was determined by the gravimetric method according to standard ISO 2060:1994. Fabric thickness was determined according to standard ISO 7211-2:1984. Standard ISO 5084:1996 describes a method for the determination of the thickness of fabric.

The tests of tensile forces of fabric in specific directions were done on samples of four woven fabrics of different raw material composition and different warp and weft densities woven in plain weave. Before testing all samples were conditioned under the conditions of standard atmosphere (relative air humidity 65 ± 2%, at a temperature of 20 ± 2°C). For the purposes of this testing standard samples with dimensions 350 x 50 mm, clamped in clamps of the instrument at a distance of 200 mm and exposed to uniaxial tensile stress till rupture.

The samples were cut in seven different directions: warp direction (φ = 90°), weft direction (φ = 0°), and at angles 15°, 30°, 45°, 60°, 75° to the weft. Three tests were done for each mentioned direction of force action on the fabric sample. Tensile properties of all samples were tested in accordance with standard ISO 13934-1:1999 using the strip method for measuring fabric strength on a tensile strength tester.

TEXTECHNO Statimat M Tensile Tester, shown in Figure 2, was used for tests. The Statimat M tensile tester is an automatic, microprocessor-controlled instrument operating on the principle of constant deformation speed. The following conditions were used for tests: clamp distance: 200 mm, pulling speed: 100 mm/min. Measurement results were collected and stored on the hard disk by the software package of the tensile tester.

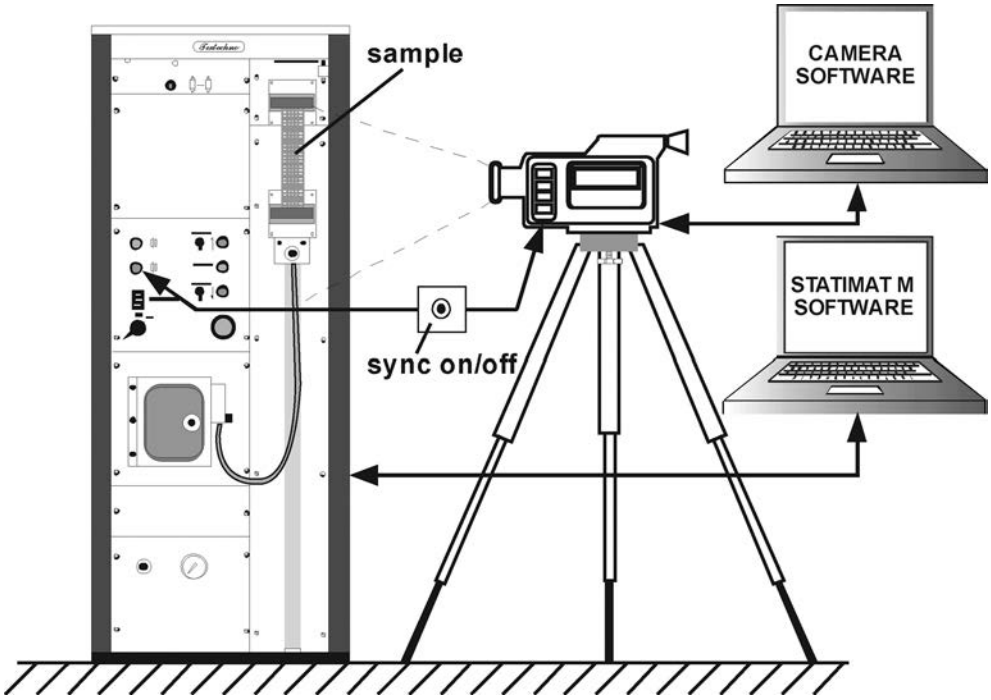


Fig. 2 – Schematic view of the experiment

3.2 Test method

For accurate recording and measurement of spatial deformation of fabric a 1x1 grid pattern was mounted on the tensile tester immediately behind the test specimen; the whole process of drawing the specimen till rupture was recorded by the Panasonic NV-GS500 Digital Video Camera placed on the tripod in front of the device as shown in Figure 2. A digital video camera with a resolution of 720 x 576 pixels, and with a recording speed of 25 frames /s, and is connected to the computer via an IEEE 1394 (FireWire) interface. The horizontal distance between camera and the sample is such that 1 mm on the grid amounted to 10 pixel. For measuring, white light is used. All the footage was stored on the computer's hard disk in MPEG-2 format. The tensile tester and the camera are connected to a special as-

sembly with simultaneous on/off which fully ensures the exactness of video recording of the entire process of stretching the fabric to rupture. The mentioned grid pattern enables fast and accurate editing of the footage processed by the software package created for this purpose which specifies the spatial deformation of samples on the basis of shifting in the direction of the x- and y-axis.

3.3 Test results

Diagrams of mean values of test results of action of tensile force F and the corresponding longitudinal strain (elongation) ε on samples when the force acts in the warp (φ = 90 °) and weft direction (φ = 0°) and under the angles 15°, 30°, 45°, 60° and 75° are shown in Figures 3-6. The related mean values of tensile force F and elongation ε of samples in the elastic range in the warp (φ = 90°) and weft direction (φ = 0°) and under the angles 15°, 30°, 45°, 60° and 75° are given in Table 2.

Table 2 – Tensile forces and elongation in the elastic range

	Cotton		Wool		Wool + lycra		PES	
φ(°)	ε(%)	F(N)	ε(%)	F(N)	ε(%)	F(N)	ε(%)	F(N)
0	1.10	5.69	1.06	6.58	7.76	0.32	2.06	1.34
15	2.06	1.32	2.98	0.69	11.04	0.22	3.66	0.76
30	6.08	1.01	5.92	0.54	11.84	0.17	5.36	0.56
45	5.56	0.73	8.50	0.41	13.40	0.17	6.46	0.39
60	5.54	1.07	6.80	0.47	9.14	0.22	5.68	0.61
75	2.94	1.60	3.94	0.88	7.22	0.31	2.84	0.75
90	1.52	3.01	1.48	3.43	3.94	0.55	1.30	1.91

From the presented diagrams, in Figures 3-6, the values of tensile force in the elastic range are used. We determined modulus of elasticity (Young's modulus) from a particular region on force – elongation curve that is determined by monitoring the experimental data obtained from an experimental set-up with regression control chart (Ozkul, Karaoglan, 2011). Using these values and using equations (2) and (3), the mean values of initial modulus of elasticity E and Poisson's ratio ν in relation to an arbitrary direction of action of tensile force on the fabric sample are calculated. The obtained experimental values of modulus of elasticity E and Poisson's ratio ν are shown in Table 3 and are used to calculate the elastic constants with respect to an arbitrary direction of action of tensile force on the fabric sample (Lo, Hu, 2002., Zheng, 2008).

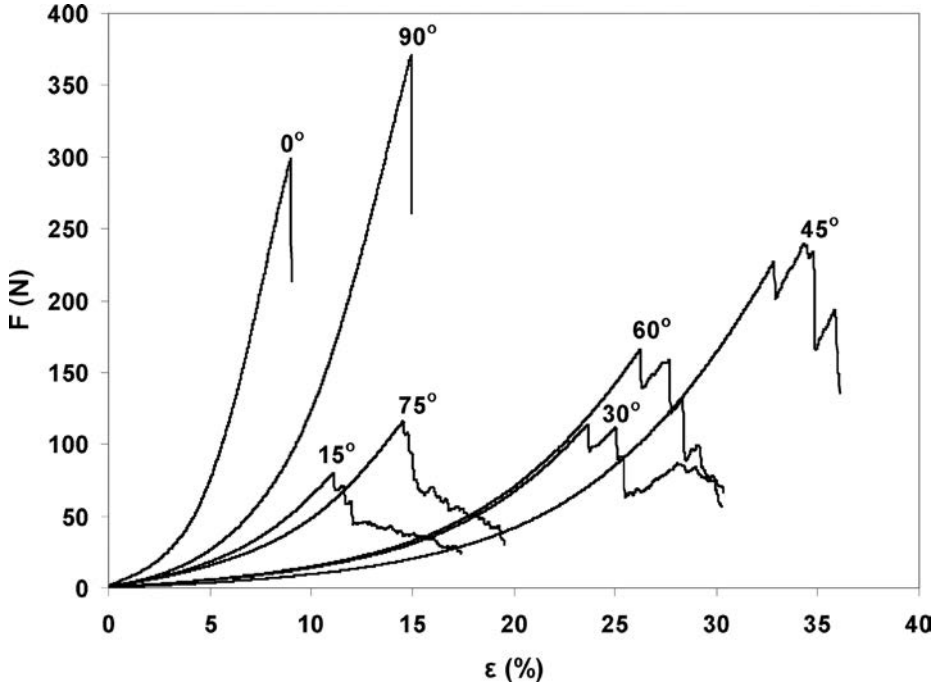


Fig. 3 – Force – elongation (F-ε) of cotton fabric

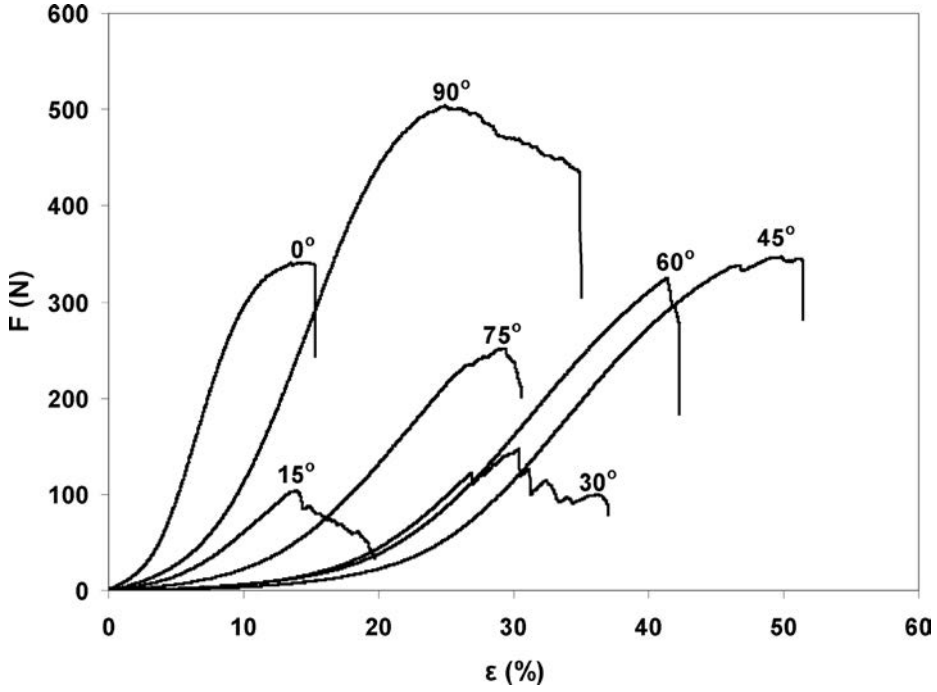


Fig. 4 – Force – elongation (F-ε) of wool fabric

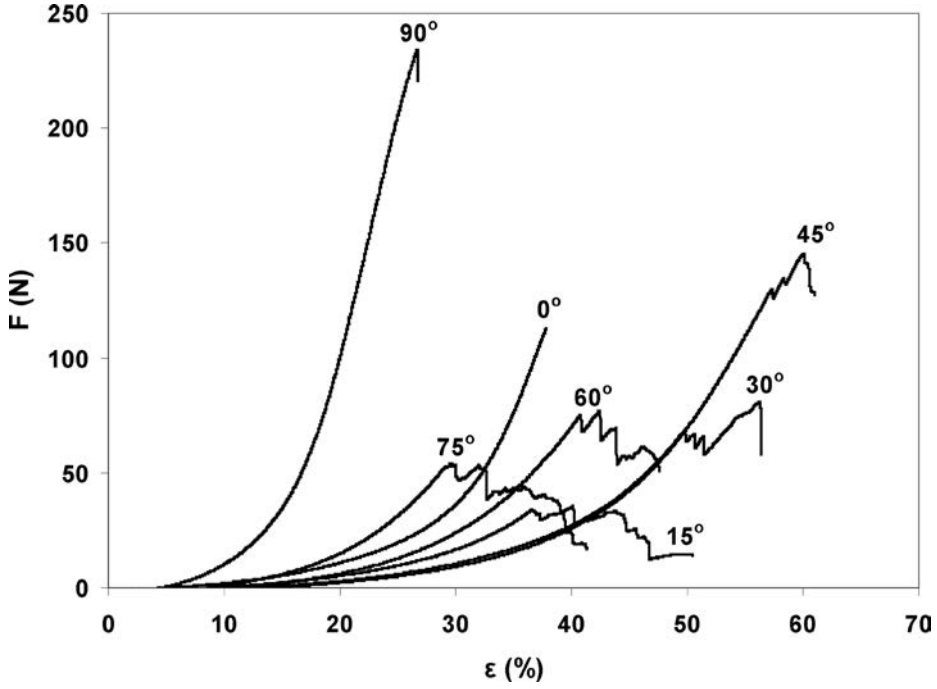


Fig. 5 – Force – elongation (F-ε) of wool + lycra fabric

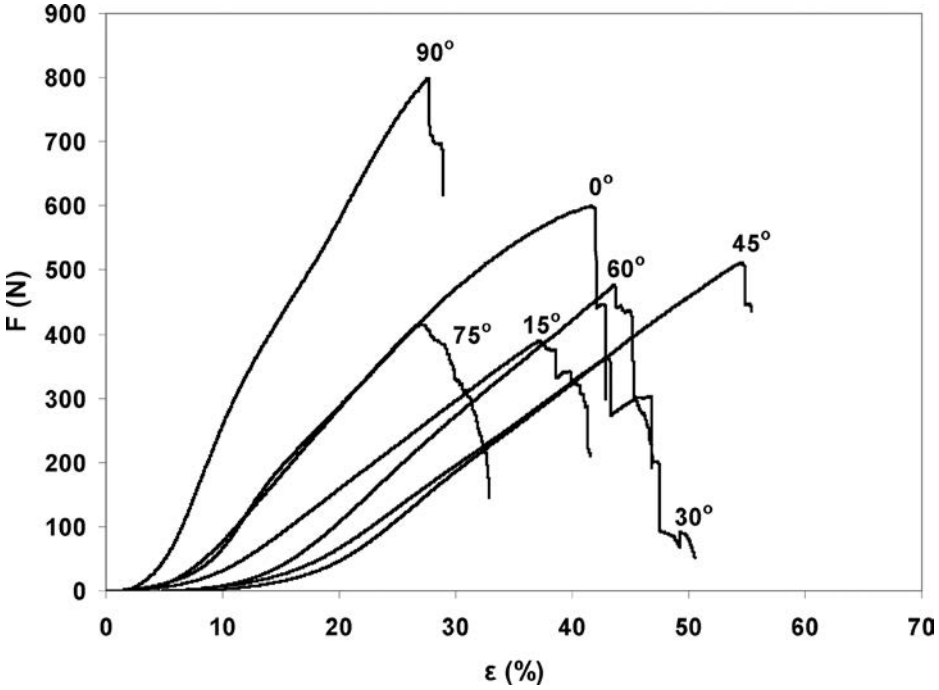


Fig. 6 – Force – elongation (F-ε) of PES fabric

Table 3 – Experimentally obtained values of modulus of elasticity (Young's modulus) E and Poisson's ratios ν

	Cotton		Wool		Wool + lycra		PES	
φ(°)	*E*[MPa]	*ν*	*E*[MPa]	*ν*	*E*[MPa]	*ν*	*E*[MPa]	*ν*
0	32.559	0.566	21.860	0.705	0.250	0.071	5.152	0.381
15	4.030	0.845	0.815	0.805	0.122	0.374	1.648	0.741
30	1.040	0.869	0.319	1.399	0.085	0.552	0.824	1.312
45	0.821	1.136	0.170	1.377	0.076	0.599	0.478	1.366
60	1.211	1.013	0.242	1.146	0.149	0.641	0.846	1.114
75	3.423	0.767	0.786	1.058	0.261	0.551	2.096	1.078
90	12.436	0.243	8.149	0.277	0.851	0.196	11.674	0.779

For each direction of force action three tests were done, and only the calculated mean values of the experiment are shown in Table 3. The tested fabrics have a greater warp density than weft density, or the same, the values E in complementary angles are not the same. Since fabrics are a special type of anisotropic materials, Poisson's ratio values fall outside the interval from 0 to 0.5, which is given for homogeneous materials such as Al, Cu, Fe, glass, etc. For the tested plain weave fabrics Poisson's ratio assumes values from 0.071 to 1.399, depending on the direction of action of tensile force are shown in Table 3.

4. Calculation of elastic constants in relation to an arbitrarily selected coordinate system

According to equations (5-11) and based on the data in Table 3 the values of elastic constants E_x, E_y, G_{xy}, ν_{xy}, α_x, α_y, were calculated, depending on a change of the action angle of tensile force in relation to the weft. The diagrams of their calculation values for each 5° are shown in Figure 7-12.

The diagrams of elastic constants (modulus of elasticity) E_x and E_y are shown in Figures 7 and 8. The lowest value E_x is always at an angle of 45° and at that angle E_y is the same as E_x. By definition E_y is equal to E_x if one of the angles (φ) is complementary to the other angle. The diagram also shows that the highest values E_x and E_y lie in the warp direction (φ = 90°) and the weft direction (φ = 0°).

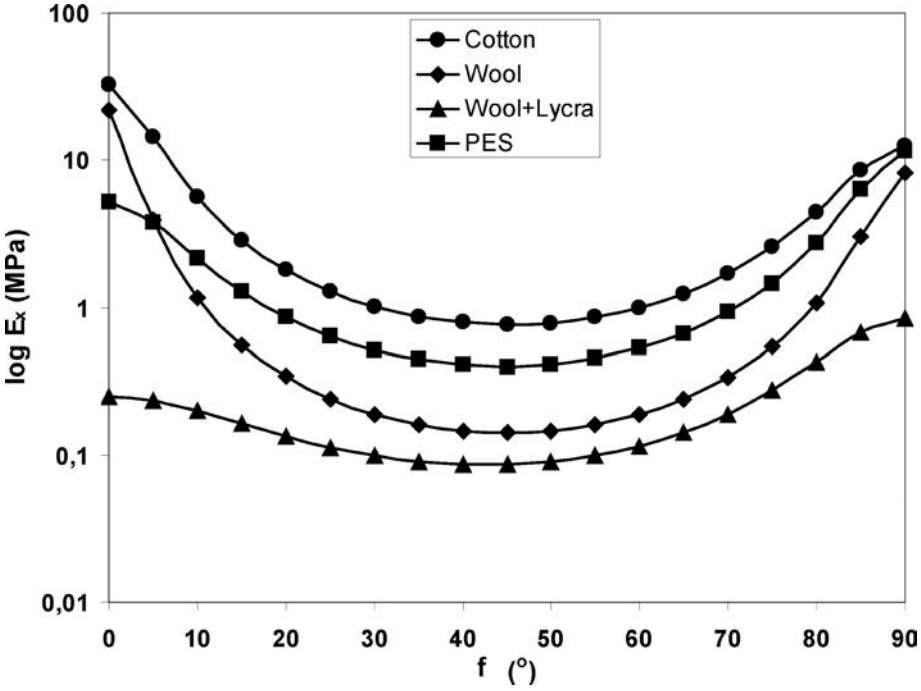


Fig. 7 – Elastic constant E_x

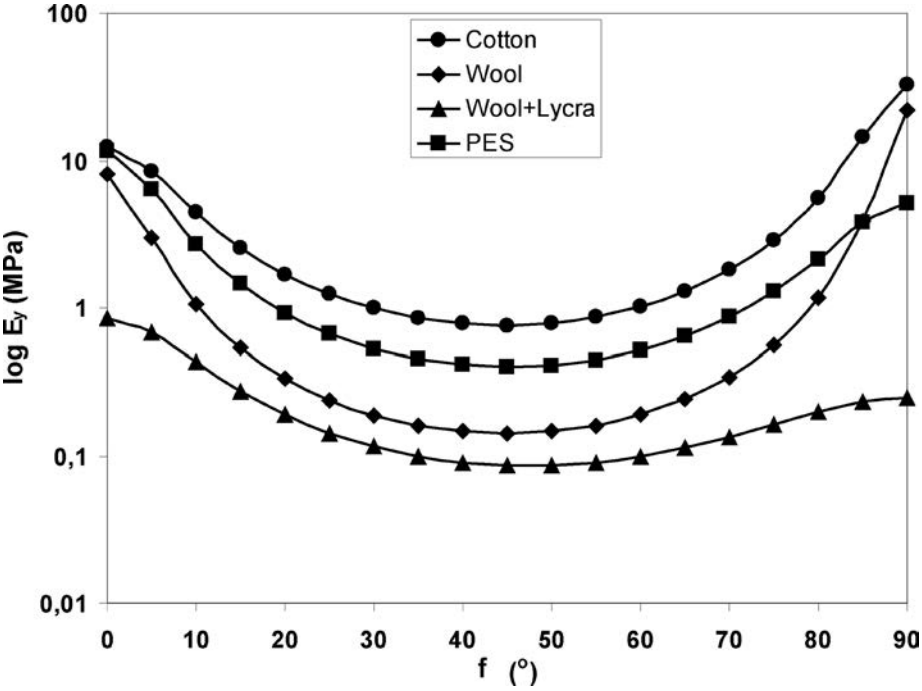


Fig. 8 – Elastic constant E_y

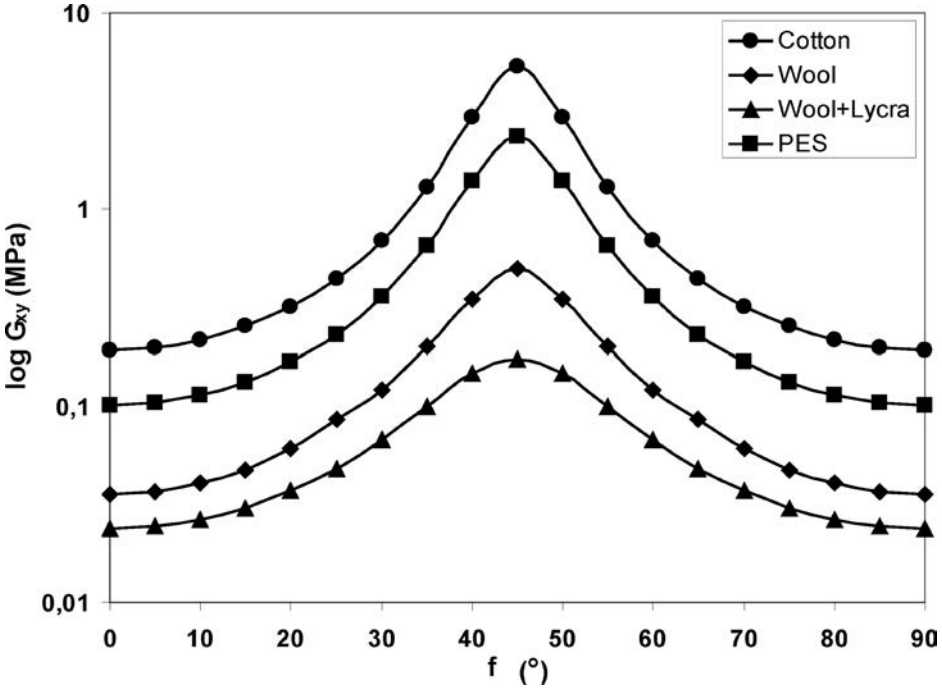


Fig. 9 – Elastic constant G_{xy}

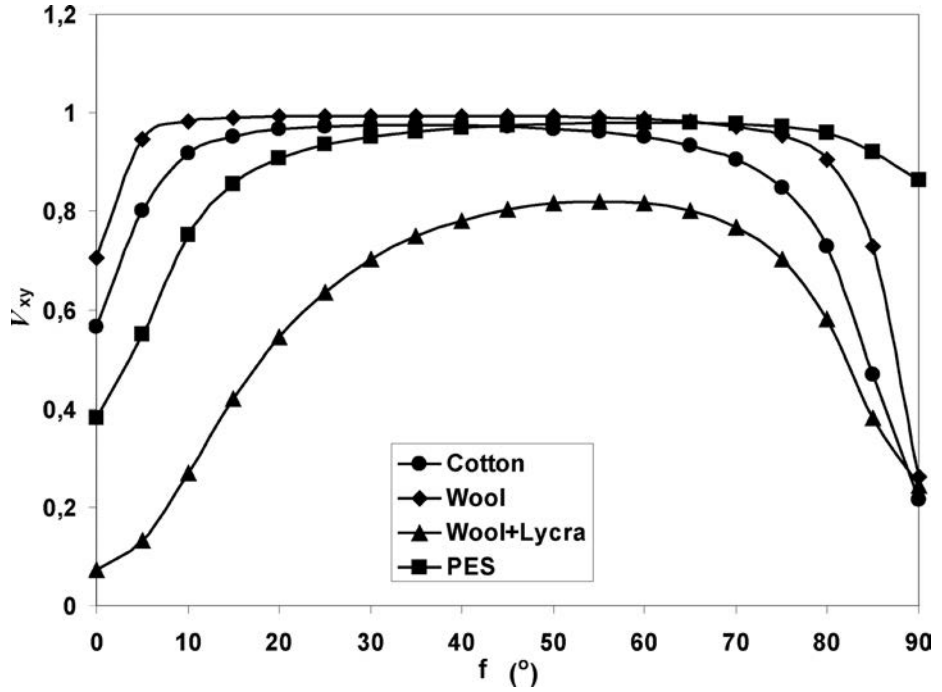


Fig. 10 – Elastic constant ν_{xy}

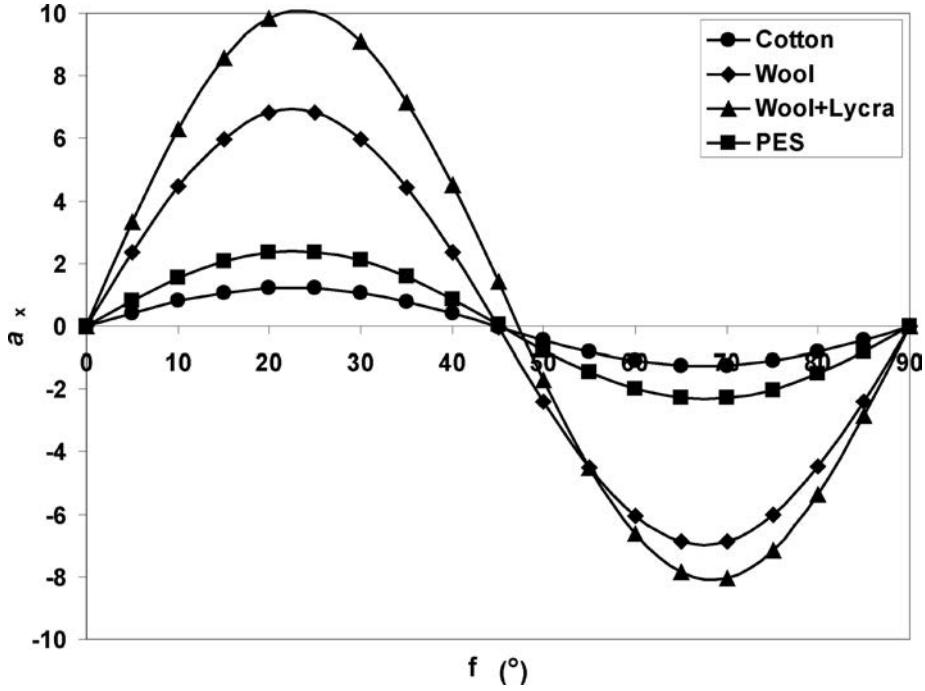


Fig. 11 – Elastic constant α_x

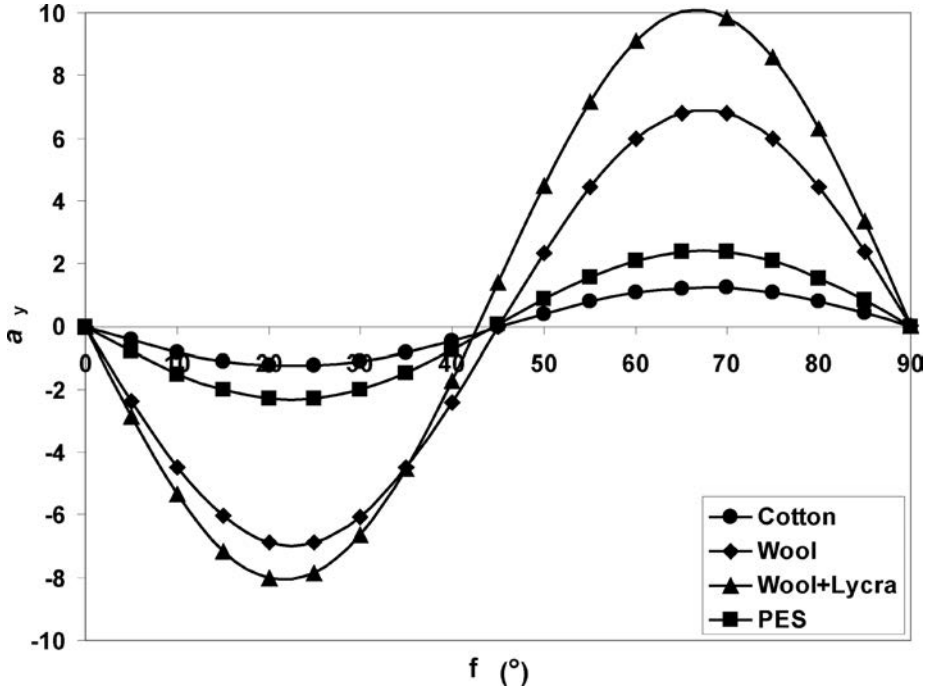


Fig. 12 – Elastic constant α_y

The values of the elastic constant E_x decrease from the weft direction ($\varphi = 0°$) seen in the angles 15° and 30°, while at the angle of 45° they assume the lowest value, and it increases towards 90° (warp direction). E_x for the cotton and wool fabric assumes the highest value when the stretching force acts in the weft direction. For the PES and wool+lycra fabric it assumes the highest value when the stretching force acts in the warp direction shown in Figure 7.

The values of the elastic constant decrease from 0° when the stretching force acts in the weft direction, at 45° they assume the lowest value and E_y increases towards 90°. Figure 8 shows the diagram E_y which assumes the highest value for the cotton and wool fabric when the stretching force acts in the warp direction, and for the PES and wool+lycra fabric it assumes the highest value when the stretching force acts in the weft direction.

The diagram of elastic constant G_{xy} (shear modulus) is shown in Figure 9. The diagram is a symmetric curve in relation to the angle of 45°. At that angle G_{xy} assumes the highest value for all fabric samples. When stretching force acts in the warp direction ($\varphi = 90°$) and weft direction ($\varphi = 0°$) elastic constants G_{xy} have the lowest value for all fabric samples. The values G_{xy} in the warp and weft direction are mutually equal for each fabric sample or it is observable that elastic constants G_{xy} are mutually equal for complementary angles.

Elastic constants E_x, E_y, G_{xy} have the highest values for the cotton fabric, followed by the PES fabric, the wool fabric, and the wool + lycra fabric has the lowest value.

The values of elastic constant (Poisson's ratio) ν_{xy} is shown in Figure 10. For the cotton, wool, wool+lycra and PES fabrics, ν_{xy} gradually increases from 0° (weft), it assumes the highest value at 45°, and then it falls when tensile force acts in the warp direction ($\varphi = 90°$). Wool+lycra fabric has the lowest ν_{xy} and wool fabric has the highest ν_{xy}.

In Figures 11 and 12 the coefficients α_x and α_y assume maximum and minimum values between the angles 0° and 90°, and their curve shape resembles the horizontally positioned letter S. It should be noted that the curve shape of the coefficient αx has mirror symmetry with respect to the curve shape of the coefficient α_y.

4.1 Comparison of the calculated and experimental results

Table 4 shows the calculated values of the elastic constants for various angles of action of the tensile force on fabric samples.

Table 4 – Calculated values of E_x and ν_{xy}

	Cotton		Wool		Wool + lycra		PES	
φ(°)	E_x[MPa]	ν_{xy}	E_x[MPa]	ν_{xy}	E_x[MPa]	ν_{xy}	E_x[MPa]	ν_{xy}
0	32.559	0.566	21.860	0.705	0.250	0.071	5.152	0.381
15	2.857	0.952	0.559	0.990	0.164	0.420	1.284	0.855
30	1.009	0.974	0.190	0.994	0.099	0.702	0.515	0.952
45	0.758	0.971	0.142	0.993	0.086	0.803	0.399	0.974
60	0.984	0.950	0.188	0.987	0.115	0.816	0.530	0.979
75	2.544	0.848	0.539	0.954	0.275	0.703	1.460	0.972
90	12.436	0.216	8.149	0.263	0.851	0.244	11.674	0.864

Table 5 shows comparison in percentage between experimental values E and ν from Table 3 and calculated values E_x and ν_{xy} from Table 4.

Table 5 – Differences in (%) between experimental E, ν and calculated values E_x, ν_{xy}

	Cotton		Wool		Wool + lycra		PES	
φ(°)	E, E_x	ν, ν_{xy}	E, E_x	ν, ν_{xy}	E, E_x	ν, ν_{xy}	E, E_x	ν, ν_{xy}
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	–29.1	12.7	–31.4	22.9	34.7	12.2	–22.1	15.4
30	–3.0	12.1	–40.5	–29.0	16.2	27.2	–37.5	–27.4
45	–7.7	–14.5	–16.2	–27.9	12.7	33.9	–16.6	–28.7
60	–18.7	–6.2	–22.3	–13.9	–23.0	27.3	–37.4	–12.1
75	–25.7	10.6	–31.4	–9.8	5.2	27.7	–30.3	–9.8
90	0.0	–11.0	0.0	–5.0	0.0	24.4	0.0	11.0

In the warp (90°) and weft (0°) directions, differences in percentage between experimental values E and calculated values E_x are 0%. It follows from the equation (5) due to the periodicity of the sin and cos functions for these values. For cotton, wool, wool+lycra and PES fabrics calculated values E_x are slightly lower than the experimental values E which can be seen with a negative sign of percentage. From the equation (8) follows that the differences between experimental values ν and calculated values ν_{xy} are 0% in weft direction. Calculated values ν_{xy} are slightly higher than the experimental values ν for wool+lycra fabric. These differences are

in range from 0% to 33%. For other fabrics, cotton, wool, PES, differences in percentage take on positive and negative values.

E_y was not taken into consideration concerning correlation due to the values of trigonometric functions of complementary angels. The results of laboratory tests of elastic constants are almost equal to their calculated values, which confirm that the above mentioned theoretical equations with high accuracy can be used to calculate the elastic constants of the fabric.

5. Conclusions

Fabrics can be defined as anisotropic elastomers, if the tensile force acting on the fabric is low. Elastic constants vary depending on the angle φ (direction of action of tensile force). Elastic constants (modulus of elasticity) E_x and E_y assume the highest values when the stretching force acts at angles of 0° and of 90°, and the minimum value is reached at an angle of 45°. The elastic constant (shear modulus) G_{xy} is symmetrical to the angle of 45° and the maximum value is reached exactly at that angle. The values of the elastic constant (Poisson's ratio) ν_{xy} for the cotton, wool, wool and lycra + PES fabric gradually increases from 0° (weft) to 45°, and then decreases in the warp direction (90°).

The coefficients αx and αy assume the maximum and minimum values between the angles 0° and 90°, and their curve shape resembles the horizontally positioned letter S. If the angles under which the tensile force acts on the fabric are mutually complementary, then $E_x = E_y$ and $\alpha_x = \alpha_y$. Different materials have different values of the elastic constants and Poisson's ratio, but the shape is similar to the corresponding curves. A good agreement between experimental results and the calculated obtained values of the elastic constants was shown. The above-mentioned theoretical equations with high accuracy can be used to calculate the elastic constants of fabrics for an arbitrarily chosen direction of action of tensile force. Therefore, the measurements need to be implemented when the tensile force acting on the fabric only in the warp, weft and at an angle of 45°.

Acknowledgements

The results shown in the paper resulted from the project "Advanced Technical Textiles and Processes", code: 117-0000000-1376, Faculty of Textile Technology, University of Zagreb, Croatia, conducted with the support of the Ministry of Science, Education and Sports of the Republic of Croatia.

References

[1] Alfirević, I. (2003). Uvod u tenzore i mehaniku kontinuuma. Zagreb: Golden marketing.

[2] Bao, L., Takatera, M., Shinohara, A. (1997). Error Evaluation in Measuring the Apparent Possion's Ratios of Textile Fabrics by Uniaxial Tensile Test. Sen'i Gakkaishi Vol. 53(1) 1997, pp. 20–26.

[3] Bao, L., Takatera, M., Shinohara, A. et al. (1997). Determining the Apparent Shear Rigidity of Textile Fabrics by Uniaxial Tensile Test. Sen'i Gakkaishi Vol. 53(4) 1997, pp. 139–145.

[4] Bassett, R.J., Postle, R., Pan, N. (1999). Experiment Methods for Measuring Fabric Mechanical Properties: a Review and Analysis. Textile Research Journal Vol. 69(11) 1999, pp. 866–875.

[5] Chen, S., Ding, X., Yi, H. (2007). On the Anisotropic Tensile Behaviors of Flexible Polyvinyl Chloride-coated Fabrics. Textile Research Journal Vol. 77(6) 2007, pp. 369-374.

[6] Herman, K. (2008). Teorija elastičnosti i plastičnosti. Zagreb: Element

[7] Hu, J. (2004). Structure and Mechanics of Woven Fabrics. Cambridge: Woodhead Publishing Ltd.

[8] Jinyun, Z., Yi,L., Lam, J., Xuyong, C. (2010).The Poisson Ratio and Modulus of Elastic Knitted Fabrics. Textile Research Journal Vol. 80(18) 2010, pp. 1965-1969.

[9] Kilby, W.F. (1963). Planar Stress-strain Relationship in Woven Fabrics. Journal of The Textile Institute Vol. 54(1) 1963, pp. T9-T27.

[10] Kovar, R. (2003). Structure and Properties of Flat Textiles. Liberec: TU of Liberec

[11] Kovar, R., Gupta, B.S. (2009). Study of the Anisotropic Nature of the Rupture Properties of a Plain Woven Fabric. Textile Research Journal Vol. 79(6) 2009, pp. 506-516.

[12] Kuwazuru, O., Yoshikawa, N. (2004). Theory of Elasticity for Plain-Weave Fabrics (1st Report, New Concept of Pseudo-Continuum Model). JSME International J, Series A Vol. 47(1) 2004, pp. 17-25.

[13] Kuwazuru, O., Yoshikawa, N. (2004). Theory of Elasticity for Plain-Weave Fabrics (2nd Report, Finite Element Formulation). JSME International J, Series A Vol. 47(1) 2004, pp. 26-34.

[14] Lekhnitskii, S.G. (1981). Theory of Elasticity of an Anisotropic Elastic Body. Moscow: Mir Publishers.

[15] Lloyd, D.W., Hearle, J.W.S. (1997). An Examination of a "Wide-jaw" Test for the Determination of Fabric Poisson Ratio. Journal of The Textile Institute Vol. 68(9) 1997, pp. 299–302.

[16] Lo, W.M., Hu, J.L. (2002). Shear properties of woven fabrics in various directions. Textile Research Journal Vol. 72(5) 2002, pp. 383-390.

[17] Ozkul, B., Karaoglan, D. (2011). Regression control chart for determination of Young's modulus: A case study. Scientific Research and Essays Vol. 6(30) 2011, pp. 6393-6403.

[18] Pan, N., Yoon, M.Y. (1996). Structural Anisotropy, Failure Criterion, and Shear Strength of Woven Fabrics. Textile Research Journal Vol. 66(4) 1996, pp. 238-244.

[19] Peirce, F.T. (1937). The geometry of cloth structure. Journal of The Textile Institute Vol. 28(3) 1937, pp. T45-T96.

[20] Penava, Ž., Šimić D. (2012). Analysis of the elastic constants of woven fabrics for at random chosen extension directions. Tekstil Vol. 61(7-12) 2012, pp. 169–179.

[21] Postle, R., Carnaby, G.A., Jong, S. (1988). The Mechanics of Wool Structures. Chichester: Ellis Horwood Limited Publishers.

[22] Sengupta, A.K., De, D., Sarkar, B.P. (1972). Anisotropy in Some Mechanical Properties of Woven Fabrics. Textile Research Journal Vol. 42(5) 1072, pp. 268–271.

[23] Shanahan, W.J., Lloyd, D.W., Hearle, J.W.S. (1978). Characterizing the Elastic Behavior of Textile Fabrics in Complex Deformations. Textile Research Journal Vol. 48(9) 1978, pp. 495–505.

[24] Sun, H. (2005). On the Poisson's ratios of a woven fabric. Composite Structures Vol. 68(4) 2005, pp. 505-510.

[25] Zheng, J. (2008). Measuring Technology of the Anisotropic Tensile Properties of Woven Fabrics. Textile Research Journal Vol. 78(12) 2008, pp. 1116–1123.

[26] Zouari, R., Amar, S.B., Dogui, A. (2010). Experimental and numerical analyses of fabric off-axes tensile test. Journal of The Textile Institute Vol. 10(1) 2010, pp. 58–68.

doi: 10.5244/C.29.155

The Self-Equalizing De Bruijn Sequence for 3D Profilometry*

Tomislav Petković[**], **Tomislav Pribanić**[+], **and Matea Đonlić**[**]

University of Zagreb, Faculty of Electrical Engineering and Computing

Using color in 3D profilometry usually requires a tedious color calibration to mitigate the undesired effects of ambient lighting, object albedo, non-equal channel gains, and channel crosstalk. We propose a novel De Bruijn sequence for multi-channel structured light that removes the need for color calibration of a camera-projector pair. The proposed sequence has the following desirable properties: (1) it enables the extraction of ambient lighting, (2) it enables the cancellation of object albedo, and (3) it enables the equalization of channel gains.

Key words:

color structured light, 3D profilometry, phase shifting, De Bruijn sequence

1. Introduction

Three-dimensional non-invasive surface profilometry using structured light is a widely studied topic in the field of computer vision with applications ranging from industrial inspection to human-computer interaction. Almost all current approaches use a digital projector to project a structured light pattern and a digital camera to record it. One of the most important issues of all such system is the design of the structured light pattern: the pattern must enable reconstruction robust to the ambient and to the characteristics of the observed object.

---

* The article was originally published in the "Proceedings of the British Machine Vision Conference (BMVC)", pages 155.1-155.11., September 2015., by the publisher BMVA Press. The article has been approved for scholarly and non-commercial use. The reprint permission granted by the authors.

** tomislav.petkovic.jr@fer.hr, tomislav.pribanic@fer.hr, matea.donlic@fer.hr

Structured light patterns may be classified in various ways (Salvi et al., 2010). The simplest classification distinguishes *one-shot* patterns that enable reconstruction from a single image and *multi-shot* patterns that require more than one image. There is usually a resolution/time trade-off involved: one-shot patterns yield a sparser reconstruction than multi-shot ones, but have a shorter acquisition time. Today a well-designed pattern should enable both one- and multi-shot reconstruction. Such patterns were proposed by Li Zhang et al. (2002) for color light and Yueyi Zhang et al. (2014) for monochromatic light.

In this work we focus on color structured light. Color structured light patterns described in the literature include fringe patterns (Wust & Capson, 1991), colored grids (Salvi et al., 1998), *N*-ary color codings (Caspi et al., 1998), and De Bruijn sequences (Pagès et al., 2005; Zhang et al., 2002; Zhou et al., 2012). The most important work by Caspi et al. (1998) describes the color imaging model that captures all the major problems of using color: ambient lighting, object albedo, non-equal channel gains, and channel cross-talk. Almost all of these problems are mitigated via the color calibration procedure, e.g. of Caspi et al. (1998) or of Juang and Majumder (2007), that must be performed prior to every imaging to include the object albedo. If calibration is performed once then the object albedo is removed by an adaptive channel equalization scheme (Zhou et al., 2012) or by an adaptive color fitting (Fechteler & Eisert, 2008), however, such approaches require either object segmentation or assumption of a constant albedo in the scene.

In this article we propose a novel De Bruijn sequence for color (or multi-channel) structured light which we call the *self-equalizing* sequence. The sequence has several desirable properties: (1) it enables the extraction of ambient lighting, (2) it enables the cancellation of object albedo, and (3) it enables the equalization of channel gains. The proposed sequence thus effectively removes the need for a precise color calibration of the imaging system, requiring only geometric calibration. Additionally, the proposed sequence enables a dense multi-shot reconstruction for static objects and a sparse one-shot reconstruction for moving objects. To the best of our knowledge none of the previous works attempted to design such a pattern. Most of the previous works use a non-altered De Bruijn sequence or impose a simple additional constraint on adjacent colors only. Our approach defines a novel constraint on the entire De Bruijn window in such way to enable the aforementioned properties. The proposed sequence shares one characteristic with the sequence of (Fernandez & Salvi, 2013): it has the same V channel, but is otherwise different.

2. Structured Light Pattern

We propose a novel cyclic color structured light pattern of which *any frame* may be used for one-shot reconstruction and of which *any consecutive m* frames may be

used for multishot reconstruction. The pattern is based on a generalized De Bruijn sequence on which we impose additional constraints that enable the robust estimation of the projected color.

2.1 Self-Equalizing De Bruijn Sequence

A k-ary De Bruijn sequence of order is a cyclic sequence of length over an alphabet of k symbols in which every subsequence of length $L = k^n$, called a window, appears exactly once on the cycle (Ruskey, 2003). A De Bruijn sequence is associated with Eulerian cycle of a De Bruijn digraph which may be used to generate it. The De Bruijn digraph contains k^{n-1} vertexes and k^n directed edges, where an outgoing edge labeled d_e (here d is a symbol from the alphabet) connects a vertex labeled $d_1d_2\ldots d_{n-1}$ to a vertex $d_2\ldots d_{n-1}d_e$, therefore concatenating vertex and edge labels yields a window $W = d_1d_2\ldots d_l\ldots d_{n-1}d_e$ of length n.

Let c be the number of channels. A full De Bruijn digraph is constructed over an alphabet of $k = 2^c - 2$ symbols. Every symbol d is represented by a c-digit binary number where value of jth digit indicates if th channel is on (1) or off (0). Note we disallow black and white where all digits of d are the same. We propose the following two constraints that every De Bruijn window W of length n must satisfy: (a) the jth digit should attain value 0 for at least one d_l in W and (b) the jth digit should attain value 1 for at least one d_l in W. These constraints ensure that all channels span the full available dynamic range, and are sufficient for the robust estimation of the projected color (see Section 3). We call a De Bruijn sequence satisfying the two constraints *self-equalizing*. The proposed sequence is constructed as Eulerian cycle in the pruned De Bruijn graph where edges belonging to the invalid windows are removed.

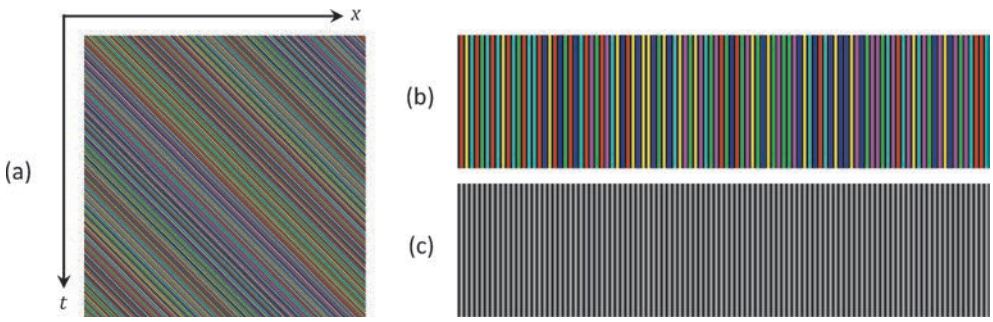


Fig. 1 – Structured light pattern for $c = 3$, $k = 6$, and $n = 3$ that contains $L = 102$ stripes. (a) shows how the pattern defined by Eq. (1) changes over x and t. (b) shows the first image in the sequence for $t = 0$; it is the first row in (a). (c) is the V channel of (b).

2.2 Cyclic Sequence of Images

The full length light pattern is a cyclic sequence of images whose cycle is comprised of $L \cdot S$ frames, of which any one frame is sufficient for one-shot reconstruction, and of which any $m = n \cdot S$ consecutive frames are sufficient for multi-shot reconstruction. $S \geq 3$ is the number of phase shifts per symbol. Let $d_1 d_2 \ldots d_L$ be a sequence satisfying the aforementioned constraints. The intensity of the ith channel of the frame at time $t \in \mathbf{Z}$ is

$$I(x,t,i) = I_{i,\max} \sum_{l=0}^{l-1} \mathbf{1}_i(d_{l+1}) \left(\frac{1}{2} - \frac{1}{2} \cos\left(2\pi \frac{x - lP}{P} - \varphi_t \right) \right) \operatorname{rect}\left(\frac{\tilde{x}}{P} - \frac{1}{2} \right), \quad (1)$$

where x is the image column index (image width must be greater than $L \cdot P$, all rows are the same), P is the period of the fringe, $\varphi_t = 2\pi\, t/S$ is the phase shift, and $I_{i,\max}$ is the maximal channel intensity. The variable $\tilde{x} = x - lP - \varphi_t \frac{P}{2\pi} \bmod L \cdot P$ ensures the rectangular function is periodically extended. The function $\mathbf{1}_i$ is the indicator function of ith channel, i.e. $\mathbf{1}_i(d_l) = 1$ if d_l contains the channel i and 0 otherwise.

We give an example for the RGB color-space ($c = 3$, $k = 2^c - 2 = 6$) where the symbols of the alphabet are: red (R), green (G), yellow (Y), blue (B), magenta (M), and cyan (C). Of full 216 digraph edges after pruning using the proposed constraints exactly 102 digraph edges remain. One possible sequence of 102 symbols is RYCRGCRCYRCGRCCRBYRBGRBCRMGRMCYBRYBYYBGYBCYMGYMCGMRGBRGMYGBYGMGGMBYCBRCBYBBYMBGMMGCMRCMYCMGBMYBMGMCRRC for which the pattern is shown in Fig. 1. Note that: (a) the V channel (the maximum of all channels) is a harmonic function enabling application of phase shifting, (b) all channels are coherent in phase which supports the spatial analysis, and (c) the full-length pattern is cyclic and repeats after $L \cdot S$ frames without any discontinuities.

3. Channel Equalization

When using color structured light the projected color must be correctly identified from the data recorded by the camera. A compensation scheme for ambient lighting, object's albedo, channel gains, and channel cross-talk is necessary. It is usually achieved by applying an imaging model whose parameters are obtained during color calibration. One model for RGB color-space widely accepted in the literature is proposed by Caspi et al. (1998):

$$I_c = \underbrace{\begin{bmatrix} R \\ G \\ B \end{bmatrix}}_{I_c} = \underbrace{\begin{bmatrix} a_{RR} & a_{RG} & a_{RB} \\ a_{GR} & a_{GG} & a_{GB} \\ a_{BR} & a_{BG} & a_{BB} \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} k_R & 0 & 0 \\ 0 & k_G & 0 \\ 0 & 0 & k_B \end{bmatrix}}_{K} \cdot f\left(\underbrace{\begin{bmatrix} r \\ g \\ b \end{bmatrix}}_{I_p} \right) + \underbrace{\begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix}}_{I_0}, \tag{2}$$

where I_c is the color recorded by the camera, A is the channel transfer matrix, K is the albedo matrix, f is a monotonic function modeling projector's non-linearity, I_p is the color instruction to the projector, and I_0 is the ambient lighting. We propose a simplified model where ith channel is modeled as

$$I_c(i) = h(i) I_p(i) + I_0(i), \tag{3}$$

i.e. the projector non-linearity f is disregarded and matrices A and K are combined into one diagonal matrix whose elements are represented via coefficient h, so the channel crosstalk is disregarded. Note that for the RGB color-space $h(1) = a_{RR}k_R$ etc. The cross-talk may be omitted if the color filters of the projector and the camera are well matched, which holds for standardized consumer electronics and which was also noted by Caspi et al. (1998). In Section 5.2 we discuss how disregarding the non-linearity does not affect spatial nor temporal processing of the proposed pattern.

Combining the model (3) with the constraints imposed on the De Bruijn sequence allows the recovery of $I_p(i)$ from I_c(i) independently for each channel. The projected value is recoverable as $I_p(i) = (I_c(i) - I_0(i))/h(i)$, where $I_0(i)$ and $h(i)$ are estimated from the recorded frame(s) using either spatial or temporal analysis.

For the spatial analysis only one frame is used. Parameters $I_0(i)$ and $h(i)$ are recovered by considering spatially adjacent stripes that form one window W. Let $m(i)$ be the minimal and let $M(i)$ be the maximal value per channel at the peak positions of n adjacent stripes. Under the assumption of spatial invariance ($I_0(i)$ and $h(i)$ are the same for n adjacent stripes) we have $I_0(i) = m(i)$ and $h(i) = M(i) - m(i)$ as W satisfies the imposed constraints. The procedure is illustrated in Fig. 2 for the RGB color-space and $n = 3$: a scanline orthogonal to the stripe direction is used to identify $n - 1$ stripes adjacent to the central stripe. Channel values are then extracted along the scanline at stripe peaks.

The temporal analysis is performed for every pixel independently. Parameters $I_0(i)$ and $h(i)$ are recovered by considering $n \cdot S$ consecutive frames that span one window W. Again, let $m(i)$ be the minimal and let $M(i)$ be the maximal value per channel. Under the assumption of temporal invariance (static scene) we have $I_0(i) = m(i)$ and $h(i) = M(i) - m(i)$ as W satisfies the imposed constraints.

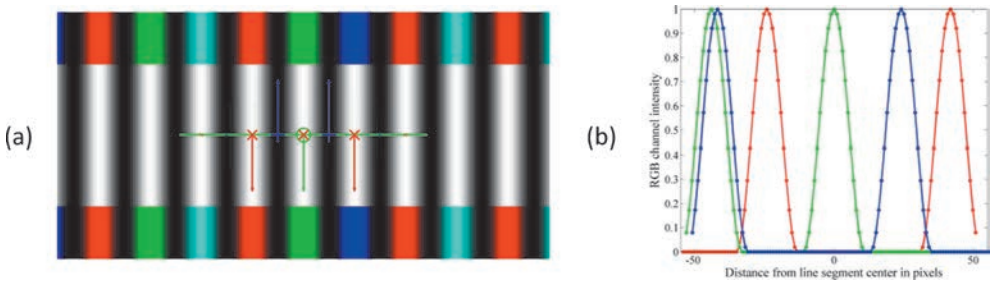


Fig. 2 – Example of spatial channel equalization for $n = 3$ and RGB color-space. (a) shows spatial position of the scanline orthogonal to the central stripe; the central part is shown in gray-scale for visualization purposes and sample points are marked with *x*. (b) shows channel values along the scanline.

Therefore, both spatial and temporal analysis enable simple channel equalization that effectively removes undesired influence of ambient lighting, object's albedo, and unequal channel gains.

4. 3D Reconstruction

As stated before *any frame* from the proposed structured light sequence may be used for the one-shot reconstruction and *any consecutive* $m = n \cdot S$ frames may be used for the multi-shot reconstruction. In this section we describe how to identify the projector light plane; once the light plane is identified the reconstruction uses the standard triangulation principle.

4.1 One-Shot Spatial Reconstruction

Single-shot spatial reconstruction is performed for every frame independently in the following steps: (a) a modified multi-scale vesselness is computed; (b) the positions of color stripes and of black slits are extracted; (c) scanlines are placed orthogonally to every detected stripe and $n - 1$ adjacent stripes are extracted; (d) only n-tuples that satisfy a spatial invariance condition are retained; (e) channels are equalized as described in Section 3, but only for the n-tuples identified in (d); and (f) stripe positions are decoded using the window property.

The multi-scale vesselness map is introduced by Frangi et al. (1998). We use a modified vesselness that extracts both bright (color stripes, $\nu > 0$) and dark (black slits, $\nu < 0$) ridges:

$$\nu = \text{sign}\,(-\lambda_{i,\sigma}) \max_{\sigma_{\min}<\sigma<\sigma_{\max}} \left\{ \exp\left(-\alpha \left| \frac{\lambda_{2,\sigma}}{\lambda_{1,\sigma}} \right| \right) \left(1 - \exp\left(-\beta (\lambda_{1,\sigma}^2 + \lambda_{2,\sigma}^2) \right) \right) \right\}, \quad (4)$$

where $\lambda_{1,\sigma}^2$ and $\lambda_{2,\sigma}^2$, $|\lambda_{1,\sigma}^2| > |\lambda_{2,\sigma}^2|$, are eigenvalues of the Hessian matrix at scale σ and where $\alpha, \beta \in \mathbf{R}^+$ are constants. Color stripes are found as the local maxima in ν and black slits are found as the local minima of (4). The sub-pixel positions are then determined as described in (Steger, 1998).

The vesselness map is computed using the average of all channels as input. The proposed pattern of Eq. (1) is spatially coherent as all channels are modulated by the same harmonic function; adding them together improves the visibility of stripes even prior to channel equalization. Using the *V* channel instead the average may reduce the performance as spatial equalization cannot be performed before stripes are found. Note that unlike previous works that use either horizontal or vertical scanlines we use a true 2D analysis. This is necessary as spatial equalization requires adjacent stripes in the direction orthogonal to the central stripe of the window; the eigenvector of the Hessian associated with λ_1 gives the required direction (Steger, 1998).

Once the scanline is placed orthogonally to the color stripe (see Fig. 2) the closest adjacent $n - 1$ stripes are identified. We define the *spatial invariance condition* as one where the extracted *n*-tuple fulfills the following: (a) the distances between adjacent stripes are $\sim 4\sqrt{3}(\sigma - 0.5)$ px, where σ is the scale of (4), (b) the color stripe directions are parallel, and (c) there is exactly one black slit between every two adjacent stripes.

Finally, the colors are identified using a simple thresholding with the threshold set at halfvalue of the dynamic range of every equalized channel. Every color *n*-tuple then, due to the window property, directly identifies one projector light plane.

Table 1 – Circular mean and circular deviation (in degrees) of the difference in wrapped phases for various channel gains. Measurement 1 uses auto white balance gains, measurement 2 uses true white balance gains (SpyderCUBE™ calibration), and measurements 3-9 use several manually chosen gains.

No.	1	2	3	4	5	6	7	8	9
Red	576	650	400	600	800	400	500	800	800
Blue	640	600	400	600	800	600	800	700	500
μ[°]	–1.0	–1.1	–1.0	–1.1	–1.3	–0.5	–0.6	–1.5	–1.9
σ[°]	5.9	6.2	7.6	6.0	7.0	6.0	6.3	6.7	7.8

4.2 Multi-Shot Temporal Reconstruction

Multi-shot temporal reconstruction is performed for every pixel independently in the following steps: (a) channels are equalized as described in Section 3; (b) the wrapped phase is computed from the equalized *V* channel; (c) stripes are detected using the window uniqueness property; and (d) the wrapped phase is unwrapped.

The following analysis assumes static (temporally invariant) scene. The wrapped phase is computed from any consecutive $n \cdot S$ frames starting from the frame at reference time t_0 as

$$\psi_w = \operatorname{atan2}\left(-\sum_{t=t_0}^{t_0+n\cdot S} V_{\mathrm{eq}}(t)\sin(\varphi_t),\ \sum_{t=t_0}^{t_0+n\cdot S} V_{\mathrm{eq}}(t)\cos(\varphi_t)\right), \tag{5}$$

where $V_{\mathrm{eq}}(t)$ is the equalized *V* channel at time *t*. The phase ψ_w corresponds to the temporal intensity model $V_{\mathrm{eq}}(t) = V_0 + V_1 \cos(\psi_w + \varphi_t)$ of the equalized *V* channel which achieves its maxima at $\psi_w + \varphi_t = 2k\pi$, $k \in \mathbf{Z}$. As $\varphi_t = 2\pi t/S$ the maxima occur at time steps $t_{\max}(k) = t_0 - \frac{1}{2\pi}\psi_w S + kS,\ k \in \mathbf{Z}$. Decoding De Bruijn symbols for $t_{\max}(k)$ that fall into $[t_0, t_0 + n\cdot S]$ interval yields a time-reversed window $W_R = d_n d_{n-1} \ldots d_1$ from which the phase unwrapping offset is computed. As $\psi_w \in \langle -\pi, \pi]$ either values $k = 0, 1, \ldots, n-1$ or $k = 1, 2, \ldots, n$ produce the required $t_{\max}(k) \in [t_0, t_0 + n\cdot S]$. The former solution gives the position of $W = d_1 d_2 \ldots d_n$ directly for the reference time step t_0, e.g. if *W* is at position *j* then the unwrapped phase is $\psi_u = -\psi_w + 2\pi j$. The latter solution requires an additional phase shift of 2π as we are decoding the adjacent window, so $\psi_u = -\psi_w + 2\pi(j+1)$. This additional phase shift is caused by the discontinuity of the atan2 function, but, as we use ψ_w directly, the discontinuity elegantly translates into one additional shift only for those pixels that require it. Also note that the obtained times $t_{\max}(k)$ are not in $\mathbf{Z}$ so we linearly interpolate between two closest time steps *t*.

5. Results and Discussion

All presented experiments were performed using Acer X1260 DLP projector, Point-Grey DragonFly2 DR-HICOL camera, and Fujinon HF9-HA1B lens. The resolution of both projector and camera is 1024 × 768. The camera-projector distance was ~11 cm. The working distance was ~1 m. The calibration volume was ~0.5 × 0.5 × 0.5 m. A geometric calibration of the camera and the projector was performed as described in (Zhang, 2000). The camera has a Bayer pattern that effectively reduces the spatial resolution making the sequence of Section 2.2 unusable due to too narrow stripes. Therefore, for the experiments we generated a shorter sequence

RYBRGCRGBRCRCYRCGRCBYRBYGBYBGRBGYBCRBCYBYCBRYC ($n = 3$, $P = 22$, and $S = 4$) that produces wider stripes.

We have performed three experiments: (1) the influence of channel gains, (2) the reconstruction of white and colored planar surfaces, and (3) qualitative reconstructions of several objects: fruits and vegetables, a human face, and a human hand.

5.1 Experiments

For the first experiment we tested the influence of channel gains to phase estimation of the multi-shot method. The camera allows setting the red and blue channel gains according to IIDC v1.32 specification (Register Reference for Point Grey Digital Cameras, 2013) while the green gain is fixed. We compared the wrapped phase of the proposed method and of the traditional gray-level fringe shifting for several noticeably off-balance channel gains and for the balanced gains. The imaged objects were fruits and vegetables shown in the first column in Fig. 4 (default channel gains), so the scene contained several different albedos. The table 1 lists circular means and circular deviations in degrees for 9 different gain pairs. Note that the mean is always close to zero and that the variance is always less than 8° (which is about one projector pixel for the center of the working area). Overall, the *self-equalization* property effectively cancels the influence of varying channel gains and of different albedos.

For the second experiment we reconstructed two planar surfaces shown in Fig. 3: a white surface and a challenging random color checkerboard. Note the perfect reconstruction for the whole white surface and for the ROI B of the checkerboard pattern. Artifacts on the right side of the checkerboard pattern were caused by the channel cross-talk that is not accounted for in the proposed model, in this case specifically from green to blue channel. Note that such artifacts are not expected for normal objects and that including the cross-talk would eliminate them. The residual error of the fitted 3D plane for the marked ROIs is 0.7463 ± 0.5483 mm for A and 0.7802 ± 0.5682 mm for B.

For the third experiment both projector and camera were used with the default factory settings simulating the most common usage scenario. The three imaged objects and the processing steps are shown in Fig. 4 where: (a) is the first of $n \cdot S$ images; (b) is temporally equalized (a); (c) is the wrapped phase computed using Eq. (5); (d) is the unwrapped phase; (e) is spatially equalized (a); (f) is textured 3D reconstruction using multi-shot method of Section 4.2; and (g) is textured 3D reconstruction using one-shot method of Section 4.1 using only image (a). Note that both temporal and spatial equalization successfully remove the ambient lightning, object albedo, and varying channel gains.

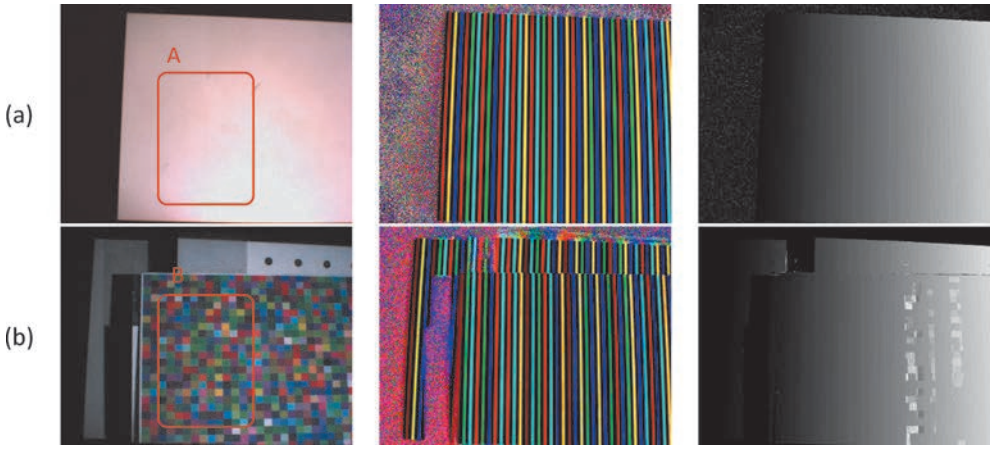


Fig. 3 – A white planar surface (a) and a planar color checkerboard (b). The first column shows camera FOV with two ROIs, the second column shows first equalized image $n \cdot S$ of images, and the third column shows the unwrapped phase of the multi-shot method.

5.2 Discussion

The performed experiments demonstrate the *self-equalization* property of the proposed pattern and showcase both one- and multi-shot 3D reconstruction from the same data. The one- shot reconstruction is spatially sparser and requires the imaged object to be sufficiently wide, e.g. for the hand image in Fig. 4 the thumb and the forefinger are partially reconstructed while the remaining fingers are lost as they are too narrow. Generally, the proposed spatial scheme is applicable only if the object is wider than $n \cdot P$ projector pixels. This is equivalent to requirement for the multi-shot method: the object must not move during any consecutive $n \cdot S$ frames. Similar constraints exists for all one-shot methods; using a projector-camera pair with a better spatial resolution enables imaging of narrower objects.

The one-shot method provided already satisfactory results even though the decoding step has been carried out directly, i.e. an additional processing step such as the dynamic programming of (Zhang et al., 2002) would further improve the results. The pattern is also compatible with the one of (Fernandez & Salvi, 2013) making a dense one-shot reconstruction an option, with a possible improvement if the windowed Fourier transform is applied to the scanlines of Section 4.1 that give the optimal window orientation and width.

The omitted projector non-linearity $f(\cdot)$ does not affect the spatial and temporal processing: $f(\cdot)$ is monotonic so the maxima and minima of the pattern always remain at same spatial position, only the perceived stripe width may change. Therefore, the spatial processing returns the true stripe centers. Phase estimation is similarly not affected as the phase of $f(\cos(\cdot))$ and $\cos(\cdot)$ is almost the same. Furthermore,

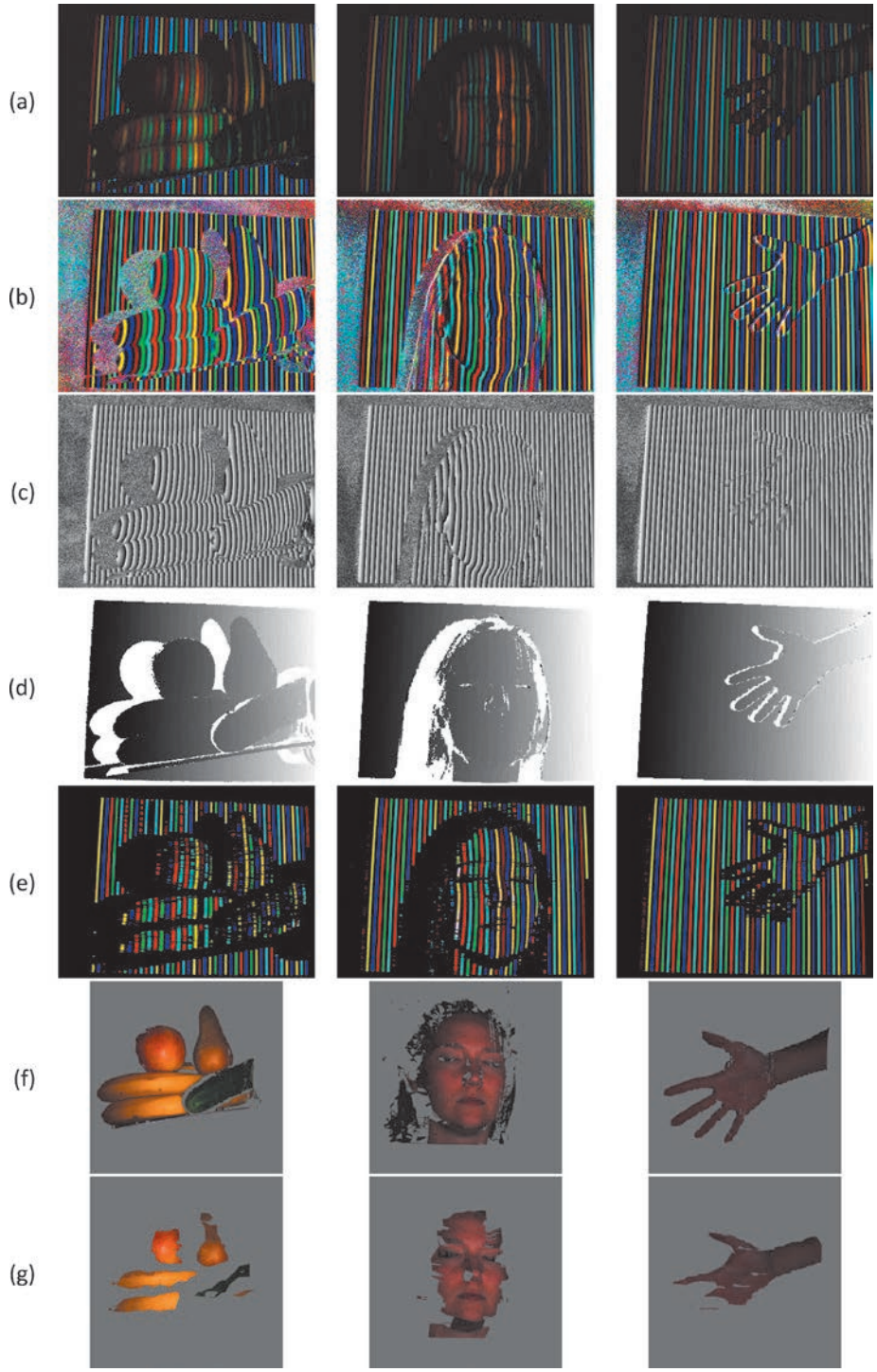


Fig. 4 – Example reconstructions for three scenes: fruits and vegetables, a face, and a hand.

the pattern is coherent in all channels which improves the spatial detection and does not affect the phase estimation. However, the coherence may increase channel cross-talk effects as the maxima of channels coincide; note that this only affects the color decoding step.

Most of the consumer grade cameras use Bayer sensors so the limiting factor for the stripe width is usually the camera and not the projector: the crest of every stripe should not be narrower than the size of the 3×3 Bayer grid. Note that the proposed method is especially suitable if a high-speed monochrome camera is paired with a DLP projector as the same color wheel having minimal or no cross-talk may be used for both.

Compared to other De Bruijn based patterns described in the literature we chose a harmonic modulating function to avoid problems with striped patterns where edges are the detected features: due to blooming effects the edge between stripes of differing brightness will move while the position of the ridge will remain unaffected.

6. Conclusions

We have proposed the *self-equalizing* De Bruijn sequence that enables time-continuous 3D profilometry where any single frame from the sequence may be used for one-shot reconstruction and where any consecutive frames may be used for multi-shot reconstruction.

We have theoretically explained and experimentally demonstrated the following desirable properties of the proposed *self-equalizing* sequence: (1) the removal of ambient lighting, (2) the removal of object albedo, and (3) the equalization of channel gains.

Future work will focus on including the channel cross-talk in two ways, through a color calibration step under the assumption of constant cross-talk coefficients, and through an additional constraint on the sequence that enables on-line recovery of all parameters of Eq. (2). Additionally, we plan to implement a movement detection in an approach similar to (Zhang et al., 2014).

Acknowledgements

This work has been fully supported by Croatian Science Foundation under Project No. IP-11-2013-3717. We are grateful to Professor Joaquim Salvi whose suggestions improved the content and the presentation of this paper.

References

[1] Caspi D. et al. (1998) Range imaging with adaptive color structured light. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20(5):470–480. ISSN 0162-8828. doi: 10.1109/34.682177.

[2] Fechteler, P. & Eisert, P. (2008) Adaptive color classification for structured light systems. In Computer Vision and Pattern Recognition Workshops, CVPRW'08. IEEE Computer Society Conference on, pages 1–7. doi: 10.1109/CVPRW.2008.4563048.

[3] Fernandez, S. & Salvi, J. (2013) One-shot absolute pattern for dense reconstruction using De Bruijn coding and windowed Fourier transform. Optics Communications, 291(0):70–78. ISSN 0030-4018. doi: 10.1016/j.optcom.2012.10.042.

[4] Frangi, A.F. et al. (1998) Multiscale vessel enhancement filtering. In William M. Wells, Alan Colchester, and Scott Delp, editors, Medical Image Computing and Computer-Assisted Intervention MICCAI'98, volume 1496 of Lecture Notes in Computer Science, pages 130–137. Springer Berlin Heidelberg. ISBN 978-3-540-65136-9. doi: 10.1007/BFb0056195.

[5] Juang, R. & Majumder, A. (2007) Photometric self-calibration of a projector-camera system. In Computer Vision and Pattern Recognition, CVPR'07. IEEE Conference on, pages 1–8. doi: 10.1109/CVPR.2007.383468.

[6] Pagès, J et al. (2005) Optimised De Bruijn patterns for one-shot shape acquisition. Image and Vision Computing, 23(8): 707–720. ISSN 0262-8856. doi: 10.1016/j.imavis.2005.05.007.

[7] Register Reference for Point Grey Digital Cameras. Point Grey, 3.1 edition, June 2013. URL https://www.ptgrey.com/support/downloads/10130/.

[8] Ruskey, F. (2003) Combinatorial generation. University of Victoria, Victoria, preliminary working draft edition.

[9] Salvi, J. et al. (1998) A robust-coded pattern projection for dynamic 3D scene measurement.

[10] Pattern Recognition Letters, 19(11):1055–1065. ISSN 0167-8655.
doi: 10.1016/S01678655(98)00085-3.

[11] Salvi, J. et al. (2010) A state of the art in structured light patterns for surface profilometry. Pattern Recognition, 43(8): 2666–2680. doi: 10.1016/j.patcog.2010.03.004.

[12] Steger, C. (1998) An unbiased detector of curvilinear structures. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20(2):113–125. ISSN 0162-8828. doi: 10.1109/34.659930.

[13] Wust, C. & Capson, D.W. (1991) Surface profile measurement using color fringe projection. Machine Vision and Applications, 4(3):193–203. ISSN 0932-8092. doi: 10.1007/BF01230201.

[14] Zhang, L. et al. (2002) Rapid shape acquisition using color structured light and multi-pass dynamic programming. In 3D Data Processing Visualization and Transmission, pages 24–37. doi: 10.1109/TDPVT.2002.1024035.

[15] Zhang, Y. et al. (2014) Real-time scalable depth sensing with hybrid structured light illumination. Image Processing, IEEE Transactions on, 23(1):97–109. ISSN 1057-7149.
doi: 10.1109/TIP.2013.2286901.

[16] Zhang, Z. (2000) A flexible new technique for camera calibration. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(11):1330–1334. ISSN 0162-8828. doi: 10.1109/34.888718.

[17] Zhou, Y. et al. (2012) Adaptive color calibration based one-shot structured light system. Sensors, 12(8):10947–10963. doi: 10.3390/s120810947.

doi: n/a

Angular Momentum Conserving Integration Scheme for Multibody System Dynamics in Lie-Group Setting

Zdravko Terze[1,], Andreas Müller[2,**], Dario Zlatar[1,**]**

[1]Department of Aeronautical Engineering, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Ivana Lučića 5, 10000 Zagreb, Croatia
[2]Institute of Robotics, JKU Johannes Kepler University, 4040 Linz, Austria

In many engineering applications, such as satellite dynamics or various case-studies of the specific locomotion patterns in mechatronics and biomechanics, motion integrals of the system need to be conserved during numerical integration in order to reflect global physical properties of the analysed motion. Derivation of such integration schemes in Lie-group settings should be especially numerically efficient since Lie-group dynamical models operate directly on $SO(3)$ rotational group, avoiding local rotation parameters, and allowing for design of structure-preserving algorithms that respect underlying manifolds of the system dynamics. In this paper, angular momentum conserving integration scheme for rigid body rotational dynamics, based on the rotational group coadjoint action, is presented and tested through the case-study of freely spinning rigid body.

Key words:

Lie-groups, Special Orthogonal Group $SO(3)$, Structure Preserving Numerical Integration Methods, Angular Momentum Conservation, Coadjoint-preserving integration scheme

[*]The article was originally published in the "Proceedings of the ECCOMAS", FAMENA, Zagreb, 2013., by the publisher Faculty of the Mechanical Engineering and Naval Architecture. The article has been approved for scholarly and non-commercial use.
[**]zdravko.terze@fsb.hr, a.mueller@jku.at, dario.zlatar@fsb.hr

1. Configuration space

The configuration space of an unconstrained multibody system (MBS) comprising k rigid bodies is modeled as a Lie-group $\mathcal{G} = \mathcal{R}^3 \times SO(3) \times ... \times \mathcal{R}^3 \times SO(3)$ (k copies of $\mathcal{R} \times SO(3)$) with the elements of the form $p = (\mathbf{x}_1, \mathbf{R}_1, ..., \mathbf{x}_k, \mathbf{R}_k)$. $\mathcal{G}$ is a Lie-group of the dimension $n = 6k$, where k is the number of the rigid bodies. The left multiplication in the group is given as $L_p : \mathcal{G} \to \mathcal{G}$, $\bar{p} \mapsto p \cdot \bar{p}$, where the product operation on $\mathcal{G}$ is defined by $p \cdot \bar{p} = (\mathbf{x}_1 + \bar{\mathbf{x}}_1, \mathbf{R}_1\bar{\mathbf{R}}_1, ..., \mathbf{x}_k + \bar{\mathbf{x}}_k, \mathbf{R}_k\bar{\mathbf{R}}_k)$ and the group identity element is $e = (\mathbf{0}_1, \mathbf{I}_1, ..., \mathbf{0}_k, \mathbf{I}_k)$.

With $\mathcal{G}$ so defined, its Lie-algebra is given as $\mathcal{g} = \mathcal{R}^3 \times so(3) \times ... \times \mathcal{R}^3 \times so(3)$ with the elements of the form $v = (\mathbf{v}_1, \acute{\mathbf{E}}_1, ..\mathbf{v}_k, \acute{\mathbf{E}}_k)$. Furthermore, the differential ('tangent map') of the L_p is defined as $L'_p : T_e\mathcal{G} \to T_p\mathcal{G}$, $(\mathbf{v}_1, \acute{\mathbf{E}}_1, ..., \mathbf{v}_k, \acute{\mathbf{E}}_k) \mapsto (\dot{\mathbf{x}}_1, \mathbf{R}_1\acute{\mathbf{E}}_1, ..., \dot{\mathbf{x}}_k, \mathbf{R}_k\acute{\mathbf{E}}_k)$ (Terze, Müller, Zlatar, 2015a).

Alternatively, configuration space of rigid MBS can be modelled as $\mathcal{G}^x = SE(3)^k$. Although $SE(3)$ represents Lie-group of proper rigid body motion, computationally both alternatives have their advantages (Mueller, Terze, 2013).

2. Dynamical model

To formulate dynamical model of the system with the direct kinematic reconstruction on $\mathcal{G}$, we introduce constrained Boltzmann-Hamel equations introduced in the form

$$\mathbf{M}(p)\dot{\mathbf{v}} + \mathbf{C}^{\mathrm{T}}(p)\boldsymbol{\lambda} = \mathbf{Q}(p, \mathbf{v}, t), \tag{1a}$$

$$\dot{p} = L'_p(v), \tag{1b}$$

$$\boldsymbol{\Phi}(p) = 0, \tag{1c}$$

where $\mathbf{M}$ is $n \times n$ dimensional generalized inertia matrix, $\mathbf{v} \in \mathcal{R}^n$, $\mathbf{v} = \left[\mathbf{v}_1, \acute{\mathbf{E}}_1, ...\mathbf{v}_k, \acute{\mathbf{E}}_k\right]^T$ are the system velocities (k bodies are assumed), while p (system 'positions') and v (system velocities with the angular velocities expressed as $\acute{\mathbf{E}}_i \in so(3)$) are defined above. $\mathbf{Q}$ represents the external and all other forces, $\boldsymbol{\lambda} \in \mathcal{R}^m$ is the vector of Lagrange multipliers and $\mathbf{C}$ is $m \times n$ dimensional constraint Jacobian, such that $\boldsymbol{\Phi}'(v) = \mathbf{C}(p)\mathbf{v}$ where $\boldsymbol{\Phi}'$ is the differential of the constraint mapping $\boldsymbol{\Phi}(p)$: $\mathcal{G} \to \mathcal{R}^m$ (Holm Darryl, 2008). With (1a) dynamic equations (Lagrangian equations of the first kind) are given and (1c) represents system

kinematic constraints. The equation (1b) represents kinematic reconstruction equation that allows for determination of the system configuration p from the velocity field ν.

3. Coadjoint-preserving integration scheme

In some of the geometric schemes for MBS, integration algorithm that preserve Lie-group structure of $\mathcal{G}$ is constructed for solution of (1b), while the system dynamical equation (1a) is often discretised via 'classical' vector-space-based numerical methods. Although computationally correct, this practice does not utilize the geometrical properties of dynamics on the rotation space that can allow for the additional useful properties of the integration algorithm, such as preservation of the integrals of motion (Terze, Müller, Zlatar, 2015b). Therefore, in this paper we propose MBS geometric scheme that extends coadjoint-preserving integration method for $SO(3)$.

Indeed, rotational rigid body equation, written as Lie-Poisson system, can be given in the form of the coadjoint operator on the dual space of a Lie-algebra (Holm Darryl, 2008)

$$\dot{\mathbf{y}} = \mathrm{ad}^{*}_{\boldsymbol{\omega}}\mathbf{y} = \hat{\mathbf{y}}\,\boldsymbol{\omega}(\mathbf{y}), \tag{2}$$

where $\mathbf{y} \in so^{*}(3)$ is the body angular momentum and 'ad*' is the dual of 'ad' operator $\mathrm{ad}_{\mathbf{a}}\mathbf{b} = \hat{\mathbf{a}}\,\mathbf{b}$ (commutator in the Lie-algebra, here identified with $\mathcal{R}^3$). By knowing that the group coadjoint orbits are conserved quantities of the Lie-Poisson system, we write solution of (2) within the each integration step as (3)

$$\mathbf{y}_{n+1} = \mathrm{Ad}^{*}_{\exp(\tilde{\psi}(t))}\mathbf{y}_n, \tag{3}$$

$$\dot{\tilde{\psi}}(t) = \mathrm{dexp}^{-1}_{-\tilde{\psi}(t)}(\tilde{\boldsymbol{\omega}}(\mathbf{y})), \;\; \tilde{\psi}(0) = 0, \tag{4}$$

with $\mathrm{Ad}^{*}_{\mathbf{R}} = \mathrm{Ad}^{\mathrm{T}}_{\mathbf{R}}$, where $\tilde{\psi}(t)$ has to satisfy equation (4) (Iserles, Munthe–Kaas, Norsett, Zanna, 2000). Thus, we use vector space Runge-Kutta method of 4th order (RK-4) to integrate substitution equation (4) (Celledoni, Owren, 2003), before the angular momentum update step is performed according to (3). Furthermore, 'simultaneously' with the integration of dynamics, system kinematic reconstruction according to the equation $\dot{\mathbf{R}}(t) = \mathbf{R}(t)\tilde{\boldsymbol{\omega}}$ is performed based on its assumed step-solution in the form $\mathbf{R}_{n+1} = \mathbf{R}_n \exp(\tilde{\psi}(t))$, where $\tilde{\psi}(t)$ has to satisfy substitution equation on Lie algebra (4). Actually, both integrations, i.e. rotational dynamics

and kinematic reconstruction, can be done by the single RK-4 step followed by the respective updates $\mathbf{y}_{n+1}$ and $\mathbf{R}_{n+1}$, which made the proposed scheme particularly efficient. Since the angular momentum upgrade is based on the group coadjoint action (3), the integration algorithm fully preserves the body angular momentum.

4. Results and Discussion

As a numerical illustration of the algorithm, we consider the motion of freely spinning body (Krysl, Endres, 2005). The initial condition is body angular velocity $\boldsymbol{\omega}_0 = \left[0.45549 \quad 0.82623 \quad 0.03476\right]^T$ and inertia tensor with the diagonal elements $\mathbf{I} = \mathrm{diag}(0.9144, 1.098, 1.66)$.

A body angular velocity (expressed in the body coordinate system) and the rotation matrix $\mathbf{R} \in SO(3)$ entries along the main diagonal as well as the matrix determinant ($\mathbf{R}\mathbf{R}^T = \boldsymbol{I}$, $\det\mathbf{R} = +1$) (showing that Lie-group upgrade $\mathbf{R}_{n+1} = \mathbf{R}_n \exp(\tilde{\psi}(t))$ reconstructs the body spatial motion that respect underlying rotation manifold) are shown in the Figures 1 and 2.

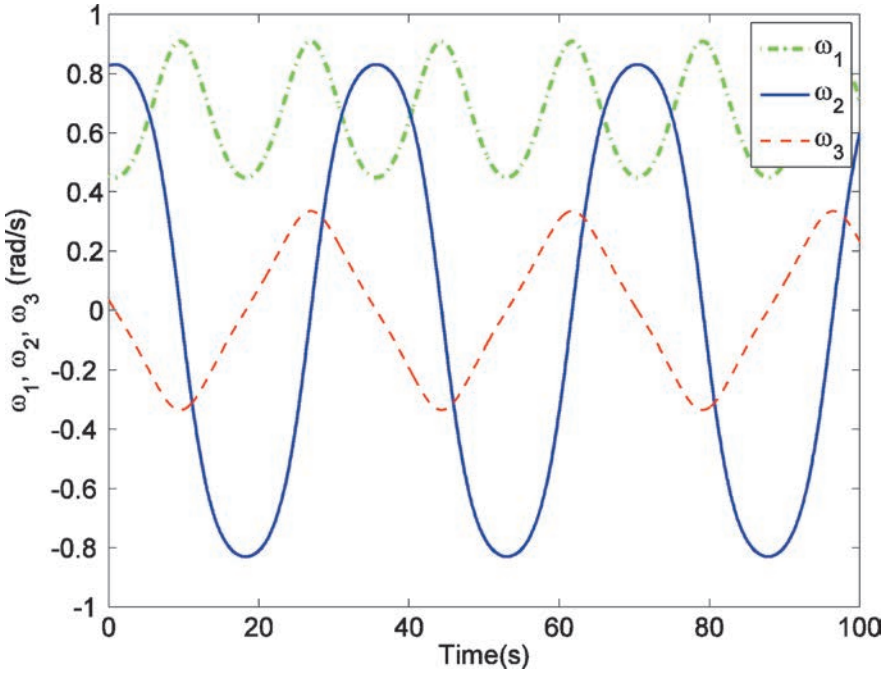


Fig. 1 – Body angular velocity.

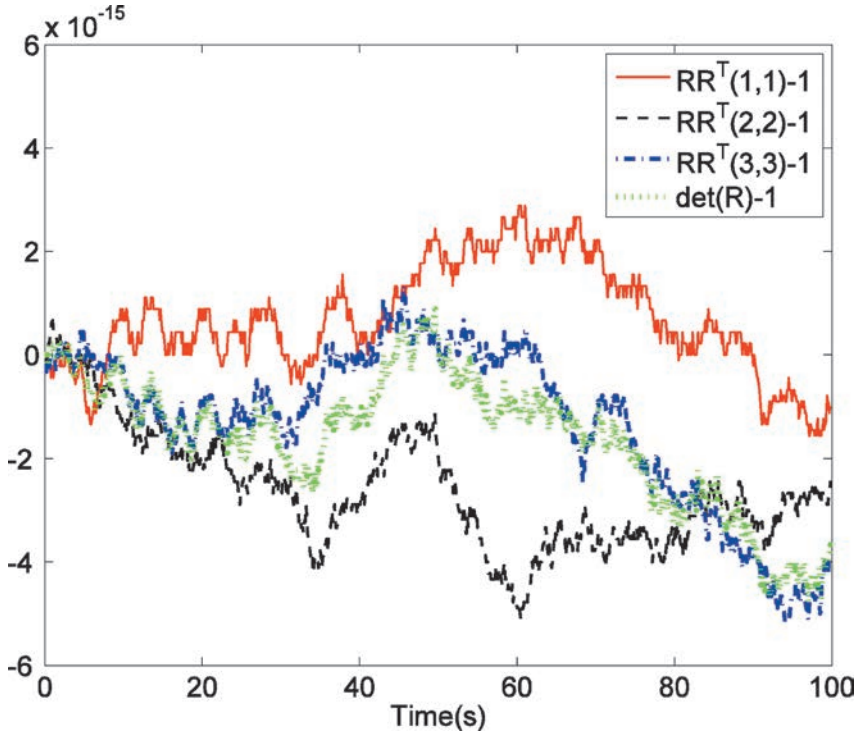


Fig. 2 – Reconstruction of the special orthogonal rotation matrix $\mathbf{R} \in SO(3)$. Diagonal elements of the product $\mathbf{R}^T\mathbf{R} = \boldsymbol{I}$ and determinant $\det \mathbf{R} = +1$.

In Figure 3, the coordinate of the body angular momentum calculated with the proposed scheme is compared to the RK-Munthe-Kaas scheme of the 4th order (where angular momentum equation is resolved in the vector space) and to the integration solution obtained with the one of the currently best second order Lie-group algorithms i.e. the method based on the Newmark vector scheme described in (Krysl, Endres, 2005). Since in the figures a drift from the constant analytical value is presented, it is visible that the proposed method yields better angular momentum preservation than the other tested numerical schemes (the other coordinates present similar comparative drift).

If, instead of integrating free-body rotation, the forcing terms have to be included, the equation (2) has the form

$$\dot{\mathbf{y}}(t) = \mathrm{ad}^*_{\boldsymbol{\omega}}\mathbf{y} + \mathbf{T}(t)\,, \tag{5}$$

where $\mathbf{T}$ is the body forcing torque. This means that we can also write

$$\dot{\mathbf{y}}(t) = \mathrm{dexp}_{-\tilde{\psi}(t)}(-\dot{\tilde{\psi}}(t))\mathbf{y}(t) + \mathbf{T}(t)\,, \tag{6}$$

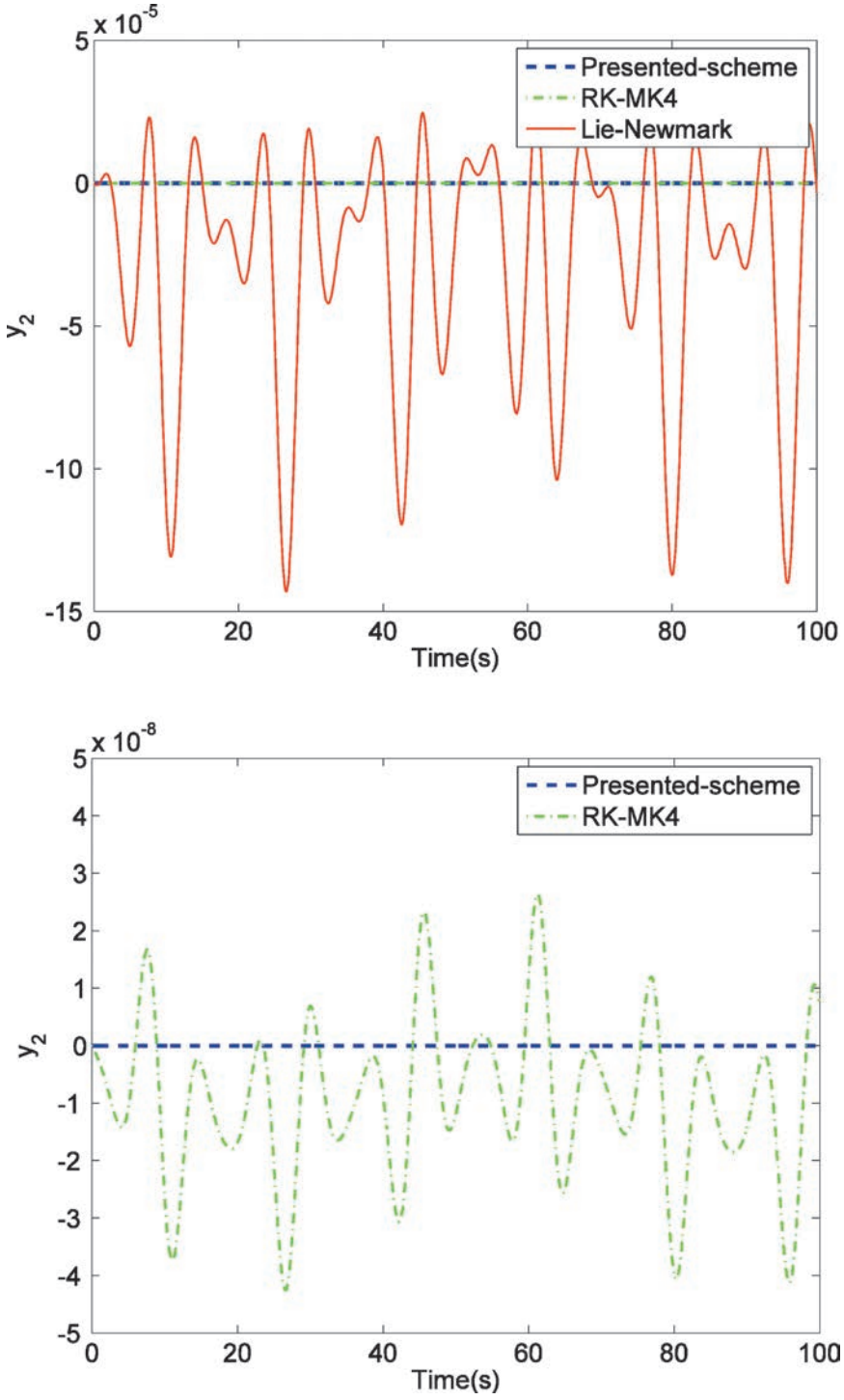


Fig. 3 – Angular momentum component drift – comparison of the integration methods.

and the angular momentum step upgrade has the form

$$\mathbf{y}_{n+1} = \mathrm{Ad}^{*}_{\exp(\tilde{\psi}(t))}(\mathbf{y}_n + \mathbf{R}_n^T \int_{t_n}^{t_{n+1}} \mathbf{R}(\tau)\mathbf{T}(\tau)d\tau), \quad (7)$$

while the RK-4 internal steps have to be modified accordingly. The described algorithm is easily incorporated in the overall MBS geometric integration scheme (1).

5. Conclusions

In the paper geometric scheme that extends coadjoint-preserving integration method for *SO*(3), and fully preserves angular momentum of the freely spinning rigid body, is presented.

Although the method is fully explicit, it generally outperforms tested semi-implicit 2nd order schemes (such as Lie-Newmark algorithm) or 4th order explicit RK-MK4 integration algorithm in terms of angular momentum preservation. In the presented form the scheme is focused on rotational rigid body dynamics (that is the most challenging part in the context of numerical integration of rigid body motion), but the method can be easily applied within the integration algorithms for dynamics of general rigid multibody systems.

References

[1] Z. Terze, A. Müller, D. Zlatar, Lie-Group Integration Method for Constrained Multibody Systems in State Space. Multibody System Dynamics 34(3), 275-305 (2015a), Springer.

[2] Z. Terze, A. Müller, D. Zlatar, An Angular Momentum and Energy Conserving Lie-Group Integration Scheme for Rigid Body Rotational Dynamics Originating from Störmer-Verlet Algorithm. Journal of Computational and Nonlinear Dynamics 10(5), 11 pages (2015b), ASME.

[3] A. Mueller, Z. Terze, On the Choice of Configuration Space for Numerical Lie Group Integration of Constrained Rigid Body Systems. Journal of Computational and Applied Mathematics 262, 3-13 (2013), Springer.

[4] D. Holm Darryl, Geometric Mechanics. Part II: Rotating, Translating and Rolling. Imperial College Press, London (2008).

[5] A. Iserles, H.Z. Munthe–Kaas, S.P. Norsett, A. Zanna, Lie-Group Methods. Acta Numerica 9, 215-365 (2000), Cambridge University Press.

[6] E. Celledoni, B. Owren, Lie Group Methods for Rigid Body Dynamics and Time Integration on Manifolds. Computer Methods in Applied mechanics and Engineering 192, 421-438 (2003), Elsevier.

[7] P. Krysl, L. Endres, Explicit Newmark/Verlet Algorithm for Time Integration of the Rotational Dynamics of Rigid Bodies. International Journal for Numerical Methods in Engineering 62, 2154-2177 (2005), John Wiley & Sons Ltd.

doi: 10.5937/grmk1404043Z

Inhabited Bridges: Art and Science (Are Architects Afraid of Designing Bridges?)*

Prof. Dr.sc. Zvonimir Žagar, d.c.**

Em. Croatian Academy of Engineering

Discussed are the development aspects and the reevaluation of some contemporary problems of the structural, civil engineering as well as architectural profession. Firstly from the viewpoint of integration of the knowledge in the IT and post-IT (IOT) era, which is crying out for basic changes in the ways of contemporary education, design methods and global perceiving of the problems which our profession should solve in surviving in the new IT environment and survival in the more and more competitive world trends. The artistic and economic approach complements the technical design aspects of inhabited bridges. There is a need to be freed of the pre-IT design heritage of the bridge design and bridge building in urban environments. The problem should be seen and solved in an interdisciplinary and multidisciplinary way.

Key words

Civil engineering, architecture, economy, rate of investments return, IT, education, bridges, inhabited bridges, synergy.

1. History of IT application in calculating bridge structures – 45 years of experience

When we (our generation) were studying the subject Bridges at the Faculty of Civil Engineering, University of Zagreb, there were no CAD tools, computers and

* The article was originally published in the journal „Building Materials and Structures“ 57 (2014) 4 (43-57) by the publisher The Materials and Structures Testing and Research Society of Serbia. The article has been approved for scholarly and non-commercial use.

** zzagar@h-1.hr

computer programs to facilitate the calculations of bridge structures (statics) and make it a rather effortless task . The drawings were literally made by using pencils and ink, triangles, T-squares, and the structures were calculated by hand, using sliding log-rulers and logarithm tables. A painstaking mind absorbing work indeed. I can proudly point to the fact that I have, then as a fellow assistant lecturer introduced the first "computer aided design", the FE structural calculation (with program STRESS) in the 1971st year, to the subject Bridges at the chair headed by the late Prof. K. Tonković. This was done, it should be admitted, with his tacit approval. Before that I did introduce the matrix method and the mentioned program for computerized calculation of structures to the students of the Interfaculty traffic and transportation study of the Zagreb University in my subject of Bearing Structures. Engs. Zdunić, Bandić (from CC Jugomont/Jugobeton) and Eng. Đurek (from ETF CC) and myself did in that 1971. year a two days crash course of the matrix method and use of the FE program STRESS, working on the old IBM 1130 installed at the ETF CC. In two "nights shift" work at the ETF CC we did all the necessary bridge structural analyses so the rest could be allocated to refining designs, finishing bridge section dimensioning and doing required drawings. It was the era of so-called "punched cards" and those green FORTRAN coding sheets. Before the modeling phase of computational model, in two afternoon hours we held a crash course from the static of finite elements (matrix calculations, which had not been taught and studied at that time), the basic of modeling of computational FE models, as well as the basics of writing data (input data) for computer based calculations. Later I have described this in detail in my handout Bearing structures I – STRESS SOLVER of engineering problems (1975). This was later followed by other more advanced FE/CAD programs: SAP, SD2A, SD2B, SAPIV and NONSAP, ICES STRUDL II, and others. Using these programs as a TOOL saves time for other CREATIVE works of study. Unfortunately, back then there was no Internet, nor WEB, no modems, not even telephone lines. And the communication with the mainframe computers (in few locations in the town) and the faculty was by foot, tram, and bus, carrying the input data and output piles mostly by foot. The output data were printed on piles of papers, and the results was drawn by hand. There were neither graphic digital displays nor plotters available. Today – this is unconceivable! But everybody was enthusiastic with it. Later on, some "exotic" desktop computers appeared and then the PC, the graphic display and finally CAD programs. In the 1980. the CAD oriented FE program COSMOS/M was obtained. Some self written programs evolve -even small structural oriented, domain oriented expert systems emerge. We installed the GURU AI expert shell and then the Wards NeuroShell – ANN shell. It should not even be mentioned that this events were crucial moments in structural design and analysis of static and dynamic (linear and nonlinear) structural behaviors. Nowadays, students are using more advanced structural analysis program packages and WEB based applets available on tablets, but unfortunately not to a desirable degree. Today, we watch 3D movies in theaters and on TV, and soon it will be unthinkable to design any structural entity without using

virtual reality (VR), and maybe, holograms too. Using these advanced software packages, students and the structural engineers, as well as architects working in practice can or will be able to focus on more creative engineering tasks – conceiving structures. These changes will be determined by investors and creative architects. All sorts of collaborate robots (IoT) will be available on construction sites. Nowadays, there is no difference between structural analyses of transatlantic liners and complex mega structures anchored in soil. Only the mega liners, cruisers, are a more complex structures, thanks to cooperation of many contributing technologies, engineering and naval skills. Not mentioning the complexity of the cruiser design, equipments and the technology involved. And not mentioning the economic and maritime skills needed.

Some remainders of historical details, as well as comparisons with the capabilities of today's IT are given in Figures 1 to 4.

Both today and historically, the key part of any bridge design was and still is the structural calculation part. By my opinion this is now just the skilled and now minor trade engineering work, mostly done by skillful use of computer programs, and skilled structural modeling. The shape of bridges was established by the exclusive need of the various transport modes (mostly railway and the road transport vehicles) required by financing clients. The today structural calculations will be re-

Fig. 1 – Example from not so faraway past (from 1956.): much is changed from that times. We may not stand still in place. Shown is the transport of a 5 Mb IBM HARD DISC 305 RAMAC HDD [24]. Now, just place your hand in your pocket and pull out the USB stick with much, much more (GB memory)!

```
      STRUCTURE OKVIR                MATOS S. MALIVUK JOVANOVIC P. BOLJKOVAC J
      TYPE PLANE FRAME               PINTAR Z. BRATIC M. ZIGIC M.
      NUMBER OF JOINTS 4
      NUMBER OF MEMBERS 4
      NUMBER OF SUPPORTS 2
      NUMBER OF LOADINGS 1
      JOINT COORDINATES
      1 X 0.0 Y 0.0
      2 X 4.90 Y 0.0
      3 X 4.90 Y -4.60  3
      4 X 0.0  Y -4.60  3
      JOINT RELEASES
      4 MOMENT Z
      3 FORCE X MOMENT Z
      MEMBER INCIDENCES
      1 1 2
      2 2 3
      3 3 4
      4 4 1
      MEMBER PROPERTIES PRISMATIC
      1 AX 0.0   IZ 0.066
      2 AX 0.32  IZ 0.02426
      3 AX 0.44  IZ 0.01100
      4 AX 0.32  IZ 0.02426
      CONSTANTS E 3000000 ALL
      LOADING 1 SVE SKUPA SA VL. TEŽ.
      MEMBER LOADS
      1 FORCE Y UNIFORM W -13.3
      1 FORCE Y CONCENTRATED P -155.6 L 1.675
      1 FORCE Y LINEAR WA 0. WB -33.50 LA 1.675 LB 3.2875
      1 FORCE Y LINEAR WA -33.50 WB 0. LA 3.2875 LB 4.90
      2 FORCE Y LINEAR WA -6.5 WB -9.60
      3 FORCE Y UNIFORM W  25.15
      4 FORCE Y LINEAR WA -9.60 WB -6.50
      TABULATE ALL
      TRACE
      SOLVE
PROBLEM CORRECTLY SPECIFIED. EXECUTION TO PROCEED.
```

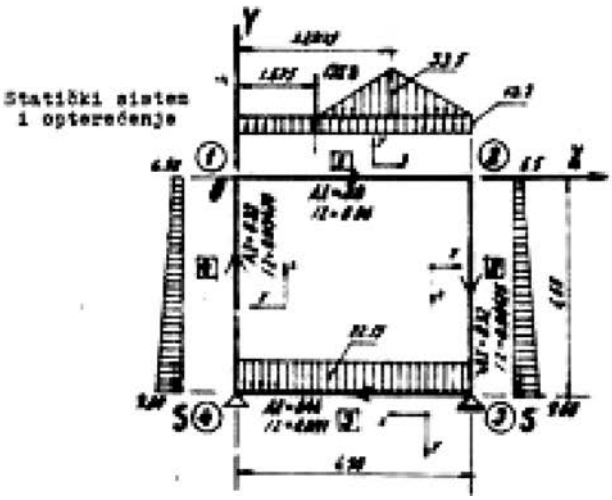


Fig. 2 – Illustration of the IT prehistory. One of the first student works describes the input data for the FE structural analyses using the STRESS program. The work was done as exercise in Bearing structures subject. The group work exercise: Matoš, Malivuk etc, students of the University Interfaculty Study of traffic and transportation engineering (in 1972.). Done on the prehistoric mainframe IBM 1130 computer at the former CC Jugomont/ Jugobeton. The input data were first prepared on the green FORTRAN input sheet, then punched on punched cards. From my handout: Bearing structures- STRESS – engineering solver, Liber, 1975. [4]

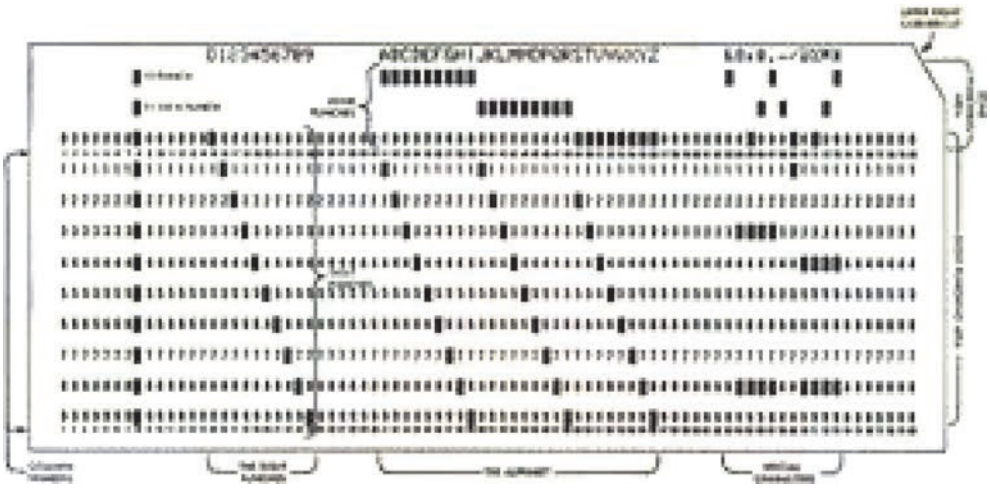


Fig. 3 – The punched card: sometimes the communication media with the mainframe computer. The prehistory of the IT era – technical archeology. [4]

placed by effective use of artificial intelligence trained expert systems. Such systems do exist even today (Adeli, Smith). The problem with modern bridge design and building lie in the complexity and the art of their execution.

During the past, the pre-IT period, the architects and urban designer were elegantly avoiding the bridge building area, excavating the tedious calculations, the dealing

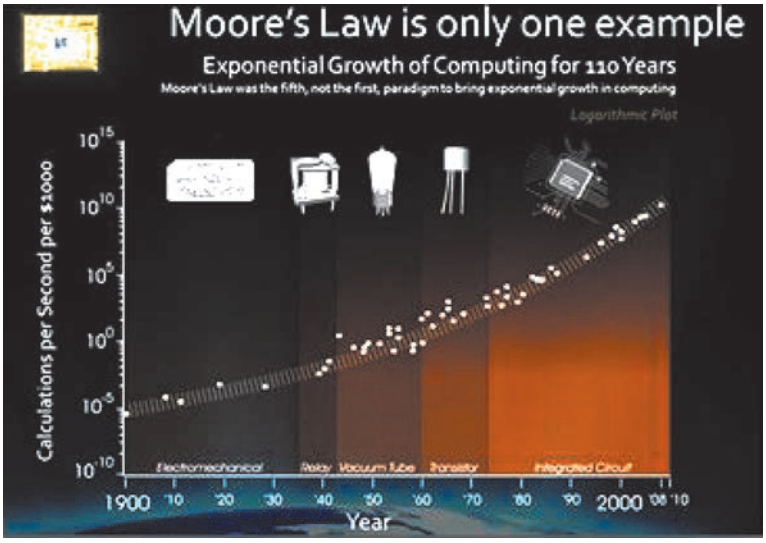


Fig. 4 – The Moors law as an example of the exponential growth of the computerization and the IT in the past 110 years. From the Ray Kurzwail speech to the GOOGLE staff delivered in 2014. (from YouTube). [15]

with new materials (concrete, reinforced concrete, prestressed concrete, prestressing, etc., then the steel and metal structures, various composite materials, glulam, CLT, etc, but mostly- dodging the tedious calculus, safety calculations, bearing capacities, deflections, vibrations, buckling phenomena, etc.), because in the bridge design were not only involved the artistic components but this "boring" work too. Here, there was no room for the art. The art was reduced and reflected in the shaping of bridge piers, the extrados and intrados curves, the "artistic shaping" of balustrades and palisades… the shape of lightning posts… and the various shaping (the esthetic) of bridge cross sections (which no observers sees)!

2. Building inhabited bridges in the urban framework

Inhabited bridges, are unlikely a novelty a new development, but probably a reflection of an era when the objects of civil engineering were conceived in an integral manner. They are known as tourist attractions, for example the Ponte Vechio in Florence, the Palladium wooden bridge in Bassano (actually several times destroyed and rebuilt), and of course the famous London Bridge, known from historical records and paintings. Also renowned, it is the inhabited bridge in Erfurt (it is the longest one, built in 1156th, and seven time burnt up and rebuilt), and there is the less famous Pont de Rohan in Landerneu (built in 1336). There are others less famous inhabited bridges too.

Recently, by the synthesis and the cooperation among architects and structural engineers emerges again, as is the artistic visions from the works of Calatrava, Zaha Hadid and others well known architects and engineers. However, over the course of history , structural engineers/bridge-builders have developed various prototypes and forms of bridge bearing structures (so known "forbilders"), heeding not much the artistic value, with no respect for art, urban needs and the needs of, mostly, professionally ignorant (illiterate) users for which bridges are actually built. The dominant question was and still is the static/dynamic analyses of the bridge structure. And the interests of various conventional modes of transportation.

It was not until recently that ideas of designing inhabited bridges have emerged as a tendency – architects are returning to this problem, as a "repugnant" numeric makes no longer a limitation to design. Nowadays manual calculations were replaced by structural modeling and procedures of finite element analysis; there are even some expert bridge- design ES. And there is CAD and recently well established VR. However, new problems are emerging, but they are related to the constructions of these very sophisticated and complex buildings, particularly mega structures. The building and design of conventional/traditional mono-modal bridg-

es has always been the the interest (the “leno”) of conventional -always the same investors/users – the rail and road vehicle transport consortiums. The cost-benefit analyses were neglected. Even with the good will, changes in these trends would be impossible to introduce in a short term, given that the education system enabled the previous (pre-IT) knowledge to be passed on from generation to generation perpetuating/repeating tasks. Besides, the very system of public tenders is “a vicious circle” because their terms are defined by professionals also educated based on the same pre-IT premises. The same applies to the nominated juries with guarantors also brought up in the pre-IT era. Thus, these bodies still lack architects who might be able to suggest new (and more complex) integral solutions, but …brought up in the spirit of a strict division between professions architects are also products of the pre-IT education era.. Therefore there are nearly no architects willing to enter such “competition”. Probably the new generations of experts brought up in the IT ere will be more prone to change in the fast approaching post-IT (IoT) era.

And why such structures would not contain various major highly profitable attractions needed from the urban inhabitants, and urban frameworks, cultural, trading, sport activities as well as other town developments. Also a profitable rational economics of investing in investments. I will not suggest any form....There are current discussions in Zagreb and in Belgrade about the need for inclusion of the rivers Sava and Danube in urban development. But those designed (as seen above) “tooth stick”- bridge-structures over rivers will never make a connection between river banks and will never integrate the complex urban towns mesh. They will just connect the various transport mode meshes only. We should strive to build inhabited bridges where the various transport/traffic modes are just one of many modes-users of the bridge space. But where are these “peculiar” objects investigated and studied nowadays? Only the derived traditional objects are studied, and these (sometimes grandiose achievements of civil engineering) are based on premises of the history of mono-modal bridges dating back from the pre-IT era with nowadays capacity of realization. In a wider sense, the integral underpass at the Zagreb railway station can be taken as a good example of an integral bridge.

It was designed by the late Prof. Tonković (IPZ) and the architect’s bureau Center 51. In fact, it is an inhabited railway-bridge built in different media (in soil), as opposed to the various mono-modal “pedestrian” underpasses, which are modeled after the sewer pipes in the ground. But we still do not recognize it as a habitable bridge, continuously paying back the invested investments. Unfortunately, this criticism towards designing conventional bridges is studied nowhere today, so it cannot promote new bridge designing tendencies. As an example of inhabited bridges let us mention here the award winning project of the famous architect Zaha Hadid (Figure 5 and 6) the first prized proposal for the New London Bridge [16], [17], [18]. For the moment, powerful lobbies have blocked its realization, as it will happen with the future proposals (if any) for inhabited bridges in Zagreb and Belgrade.

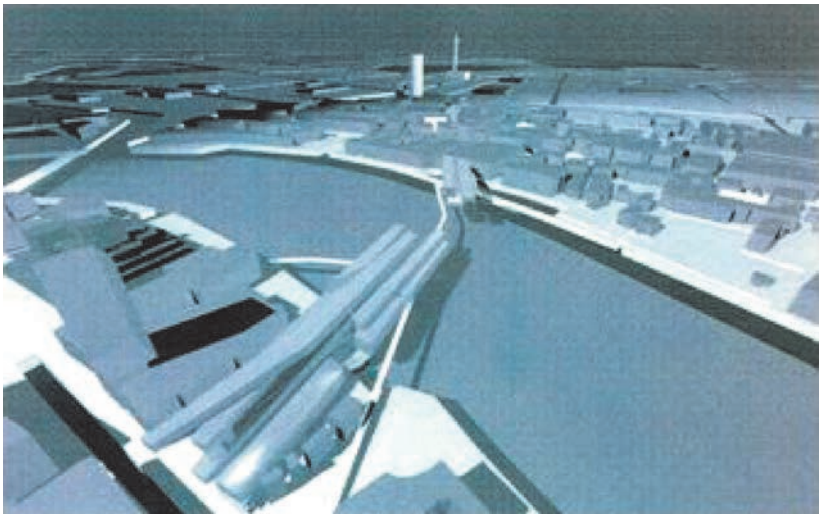


Fig. 5 – Proposal for the inhabited London Bridge(2009),by Zaha Hadid [16]

For example, when someone in Zagreb considers expanding the Remetinec (short-sightedly designed) traffic rotor by the Sava bridge, he fails to take into account the possibility of expanding it over the Sava river, the possibility of including the architects in the project, announcing an international tender, maybe resolving the issue as an inhabited object over or above the Sava – it is namely the "fief" of the traffic, construction, and completely different interest of the (hydro-technical) group.

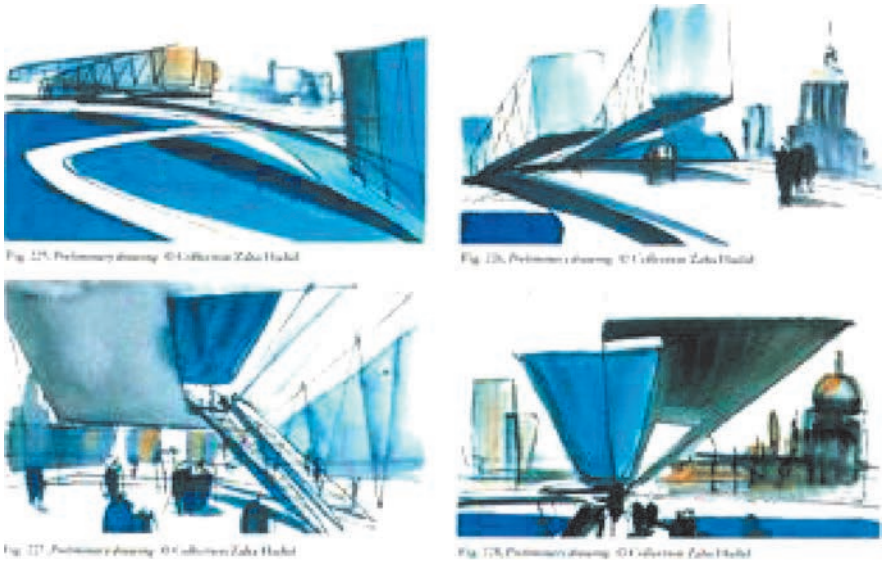


Fig. 6 – Few sketches for the inhabited New London Bridge proposal by Zaha Hadid. [17]

Also it is worth to mention the winning project of the recent competition for the new inhabited London Bridge (Inhabited London Bridge 800 Competition, RIBA 2009) designed by Laurie Chetwood Architects group (Figure 7), as well as some other proposals (Figures 8, 9) [18].

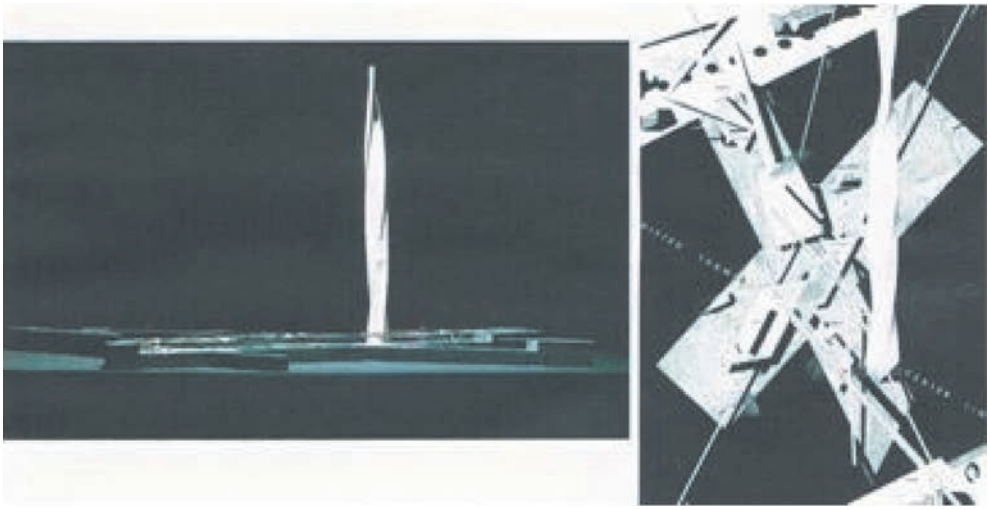


Fig. 7 – Daniel Libeskind: proposal for the inhabited bridge over Themes in London (2009) [18].

The resistance to the change is monumental. The vulnerability of various interest areas is omnipresent fear of affected interest groups, inflexible to passage of time and advancement of technology

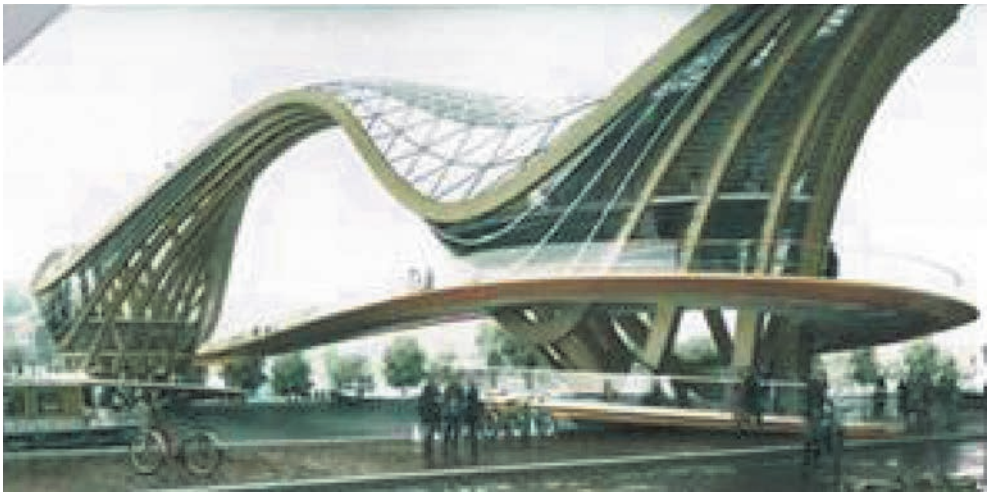


Fig. 8 – Laurel Saint-Val: proposal for an Amsterdam bridge. Some new forms [19].

Fig 9 – Pedestrian Helix-bridge in Singapore modeled after the DNK spiral. Cox Rayner Architects, [20]. Large sum of money were invested in the structural and bridge structure model research.

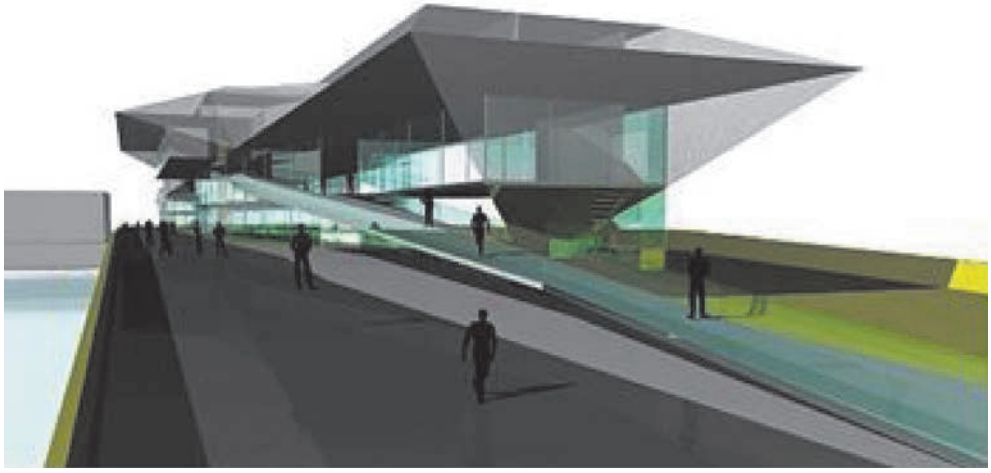


Fig. 10 – Chris van Niekerk: The Fold Architecture (2009): an idea for an inhabited bridge [18].

But, gradually as time is passing by, there will be changes, because the consciousness of the obsoleteness of pre-IT static views on the bridge design will prevail. It is not all in the structural mechanics and static as well as dynamic analysis. Radical prevailing changes have already appeared in the past history of our profession. For example, some fancy steel trussed bridge forms, formed on the basis of the "beauty" of min/max bending moment envelope diagrams and some questionable savings of steel simply disappeared from our views. The change of views will be caused by recognition that the integration of various bridge contents is beneficial to the economic health, pleasant and attractive to users, enjoyable to walk trough and profitable to investors. This will take some times, because the contemporary art of prev-

alent bridge design types are perpetuated from previous (prototype) built forms studied at bridge design faculty chairs all around world. This is a kind of a last forever "Ravels Bolero". Time and will are needed for necessary various knowledge integration as well as needed creativity. And by the way: had anyone ever invited, for example, Zaha Hadid or Antoine Grumbach or Richard Rogers or Antoine Grumbach or Richard Rogers or Yon Friedman or Daniel Libeskind to lecture or give lectures, hold seminars to students of the Civil Engineering faculties and/or the Faculties of Architecture, about their views to contemporary and future bridge designs.

+++

The time is needed to comprehend that for a bridge design the economy of investment and the management cost of maintenance is as important as are the building techniques and "conventional" structural calculations. It is important to notice that the returning of invested money (the invested capital) is of paramount importance to healthy economy investments. These econometric algorithms (developed by Aguilar) have long been known and well tested, but consciously or unconsciously (or out of ignorance or because of lack of cooperativeness) are avoided, because they threaten the "conventional" design of bridges. It should be clear that the cost of building habitable bridges is much higher than that of conventional monomodal forms. Undoubtedly, the structure of inhabited bridges is far more complex (Figures 10-12) than of conventional (mono-modal) bridges.

On the other hand, nowadays we have much better and "powerful" analytical tools available for analysis structures and investments. We do have sophisticated FE analytical tools, expert systems, for static and dynamic analyses, as well econometric packages and models and computational techniques for cost-benefit, econometric and system analyses of sophisticated complex constructions. We do have powerful CAD, VR, augmented reality systems and soon will have 3D visualization tools. Such complex structures as are inhabited bridges may not be designed by structural engineers only, these objects require close cooperation across several professions/specializations and several investment teams, a big leap forward in cooperation and understandings – a large difference from the today practice and affairs.

3. The dangers of applying knowledge acquired in the pre-IT era

Our (old) generation of civil and structural engineers, met computers relatively late in their lifetime and just recently were introduced to smart IT gadgets (iPods, tablets, smart phones, Google glasses, 3D printing and various on Web applets). To the

new generations using and manipulating those gadgets as well web, internet, playing computer games, using Google and other search engines is something natural as a clever/smart TV sets: the extensions of their nerve system. We (the old generation) are maybe using all of those gadgets, tools and programs but those tools will never be the integral part of our neural system, the extension of us and our minds. We are, for example, still suspicious to AI applications, expert systems and the use of robots in design and in building and construction. Most of us then look at those "innovations as on a kind of no desirable "witchcraft". Other (technical) professions are bypassing us by big strolls. The bigger danger to advancing the profession is generated by the current knowledge transmission from the generation to generation using the, before mentioned, perpetuating way of knowledge and skills transmissions from the past-IT and pre-IT times (sometimes even the irrelevant facts of the past) – transmitting it to the questionable future, and insisting, by the way of education, to force the past and sometimes irrelevant facts to the new student generations, who will work and deal in to us unknown IT and post-IT environment of the ever speeding IT development (for example, the pending 4th Industrial revolution and the IoT-internet of things). As well as in 2011 Udacity founded online nanodegree courses (MOOCs) in future university education. We still do resist the force by force: the action with reaction. Not even considering the possibility of intelligent reaction of bionic structures. Do we even research bionic structures in civil engineering studies?

In that permanently established "traditional" knowledge transfer, founded many centuries ago, there are still cultivated traditional (even today outdated) skills and

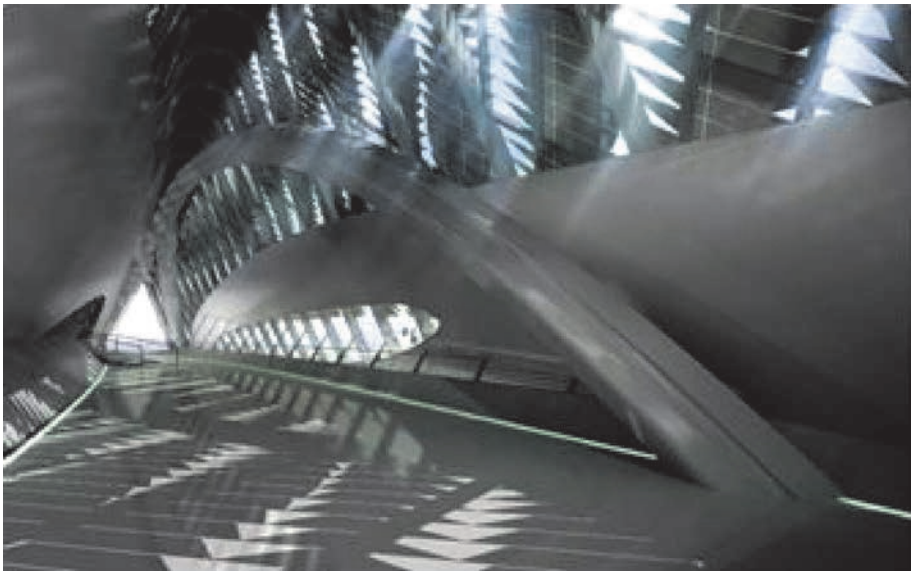


Fig. 11 – Zaha Hadid: the Zaragoza bridge pavilion. A living bridge (2008) [16].

obsolete historic facts, unessential to the contemporary and future generations of civil and structural engineers. The generation who will work in environment in which machines/computers will talk with other machines/computers (note the pending 4th industrial revolution era IoT). I was earlier proposing a radical outsourcing of those outdated skills and knowledge (so important in the past pre-IT days, but even obsolete today) to a new founded (technical or other ways named) archeology study, where those past technical/mechanical skills and methods could be archived and studied and even improved. I am convinced that this is similar in other technical domains too.

My opinion is that technical museums are not a solution to that problem, because in them are exposed dead technical artifacts and not still "living" but outdated (graphical and numerical methods, theories etc.) facts. About which numerous technical books, pamphlets, manuals, programs, scientific articles, studies and even dissertations were written and published. In that way the core of the modern technical study will be freed from the "glorious" technical, now irrelevant, pasts, historical facts and therefore the core of the studies divided/cleaned from irrelevant facts and the future engineers guided to the engineering problems in the approaching post-IT era. On the other hand we have to be conscious of the emerging new forms and teaching and study trends, the online studying on WEB based university studies and ours involvement in that. This poses the question of future (physical) developments of the universities in the view of fast developing online universities and specializations (see TED on YouTube). In those trends our universities faculties and chairs should be included. And of course the students. Without any delay the students should be guided toward timework and interfaculty cooperation cooperating with similar or inbound faculties in the same building, in the same town, outside towns, with abroad located high school institutions to gain from such cooperation. The students should be headed to projects and not to subjects; the study should be, from the first year, project oriented, instead of subject oriented (us nowadays). This requires the cooperation among faculties of the world, and well the mass exchange of students, teaching and research staff. At present times there is practically no cooperation of the Zagreb Civil Engineering faculty, the Faculty of Architecture and the Faculty of Surveying located in the same building (!), not mentioning the other faculties in the town or State. The so needed close cooperation with the FER (Faculty of Electrical Engineering and Computing) and the Faculty of Mechanical Engineering and Naval Architecture as a prerequisite for research and developing of smart and complex structures, robotic applications and energy savings are practically nonexistent. There are maybe some weak links with the Faculty for Economics. This could be seen from the titles of diploma works, the PhD titles, research papers etc. in yearly published faculty annuals. But without such cooperation among cited faculties and other large enterprises there could not be inhabited and economically healthy bridge designs. The bridge must be an economically healthy synergetic product.

4. … and instead of conclusion …

Recently I was approached by some experts in the field of architectures and economics who are researching (obviously interested thematic) of our past professional engagement and involvement into the development affairs of the development " third world" counties worldwide, in which our experts and enterprises from various fields were acting and involved. Those were fields of medicine, medicare, architecture, civil engineering, technology, education, architecture, urban development, construction, industries, trade and other technical and social fields. The foreign students were studying here at our faculties and high schools. Our experts were present and cooperating in various African, Asian and South American states, as well in the former SSSR. The thematic is obviously of a wider interest. The documentation is sparse. Researching the work of Yugoslav architects and civil as well as mechanical engineers in the non-aligned countries, is now a challenging task, since documentation is sparse, even nonexistent, since documenting the processes and completed projects was the last thing on the mind of the construction companies and then the (former) Technical aids, focused mainly on progress and growth. We have lost all those markets, links, connections, personal relations and all the business cooperations, due to our ignorance and stupid self-satisfactions. All was not in the economy crises only. Now the cited researchers are searching for the

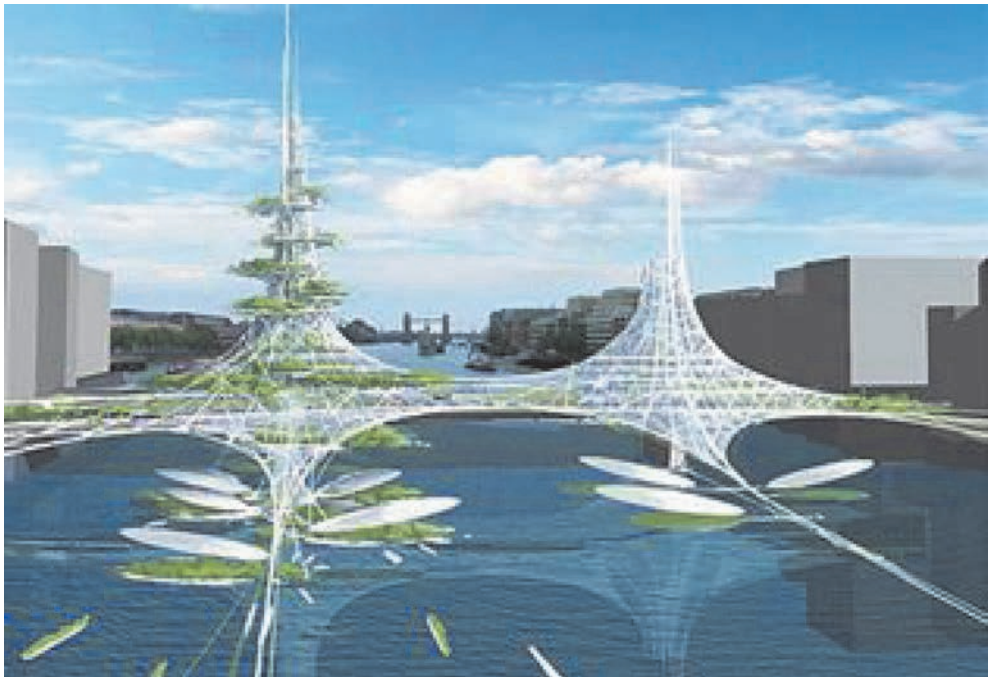


Fig. 12 – The winning work of the Inhabited London Bridge Competition (2009) a new by Laurie Chetwood, London. It is an energy self-sufficient and self- financing "green" structure [18].

Fig. 13 – Proposal of the young Florence architects group (Gruppo Giovani Architetti Firence) for a new inhabited bridge over Arno in Florence.

Fig. 14 – Stiphen Hools proposal for a inhabited bridge in Copenhagen (2008). The structure holds economic self keeping technology, wind turbines, photo volatile skyscraper casings. There are public spaces, offices, shops, terraces, various pedestrian communications, excavators, elevators and pathways. [22]

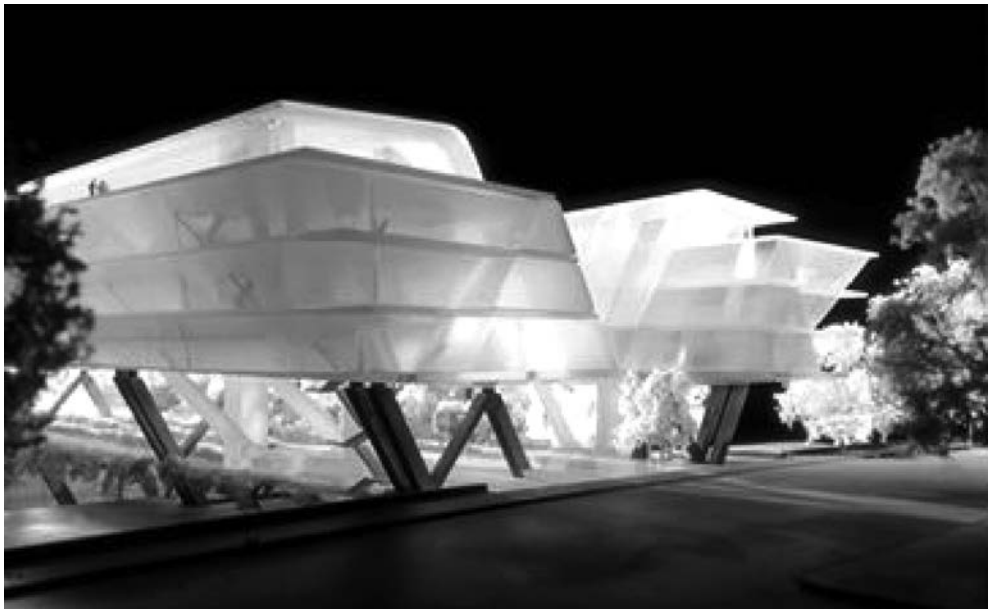


Fig. 15 – A proposal of an inhabited "garden-bridge for the 21st century" in Rome (1997), Studio Bednarski Ltd. [23]

complex mechanism of those collapses. To regain the lost cooperation and the lost markets in a worldwide competition we have to fast change our education system and include ourselves in that competition with innovations, new ideas and transforming ourselves as fast as possible.

Only applying the most recent applications, research and advancements and IT and the post-IT technologies (the IOT), as well as the most recent development trends, with the inclusion in the international work distribution markets, we may be able to be again competitive, be present and interesting in the worldwide market. Where could we better establish such (lasting) links, then in the early faculty study stage and the international (and interdisciplinary) problem oriented online cooperation of worldwide student cooperation? We should (try) to be the first – and why not?

References

[1] Z. Žagar, Nove koncepcije mostova (New concepts in bridge design), Ceste i mostovi (1973) (3) 67-78.

[2] Peter Murray, Mary Anne Stevens: Living bridges, Prestel Verlag, Munich, 1996.

[3] R.J. Aguilar, System analysis and design in engineering, architecture, construction and planning; Prentice-Hall, 1973.

[4] Zvonimir Žagar, Nosive strukture, I dio; 2. Bearing Structures, 2nd Ed, Zagreb University Press, u Zagrebu, Faculty of Civil Engineering, Liber, Zagreb, 1979.

[5] Zvonimir Žagar, Drvene konstrukcije I (treće izdanje), Wooden structures 1 (3rd Ed.), Pretei, Zagreb, 2002.

[6] Zvonimir Žagar, Drvene konstrukcije II (treće izdanje), Wooden structures 2 (3rd Ed), Pretei, Zagreb, 2002.

[7] Zvonimir Žagar, Drveni mostovi, (drugo izdanje), Timber bridges (2nd Ed), Pretei, Zagreb, 2006.

[8] Zvonimir Žagar, Kako nadvladati zasade naslijeđa u školovanju građevinskih inženjera (I dio), How to overcome the inherited roots in the education of civil engineers (Part 1), Tehnika, Naše građevinarstvo, year 60-2006, No.4, pp. 8-14.

[9] Zvonimir Žagar, Kako nadvladati zasade naslijeđa u školovanju građevinskih inženjera (II dio), How to overcome the inherited roots in the education of civil engineers (Part 2), Tehnika, Naše građevinarstvo, year 60-2006, No.5, pp. 13-18.

[10] Z. Žagar, Zagreb i nad vodom, (Zagreb above the Sava river),Tehnika – Naše građevinarstvo, 2009, vol 63, No. 3, pp. 9-16.

[11] Dubravka Sekulić: Tri tačke oslonca: – three points of support: Zoran Bojović, Muzej suvremene umetnosti, Contemporary Art Museum, Beograd, 2013.

[12] Recommendations for implementing the strategic initiative INDUSTRIE 4.0, Federal Ministry of Education and Research, Forschungsunion i Acatech, National Academy of Science and Engineering, Germany, April 2013.

[13] Mišljenje o informacijskoj i komunikacijskoj tehnologiji u Nacrtu prijedloga industrijske strategije RH 2014.-2020.(temeljeno na zaključcima okruglog stola održanog 27. veljače 2014. godine), Znanstveno vijeće za tehnološki razvoj HAZU, Opinion of the Academy on the Proposal of the indusrial strategy development of Croatia papers, Zagreb, 2014.

[14] Smart Manufacturing Coalition-led Project Wins DOE Clean Energy Manufacturing Contract, @Nimbia Services, SMLC – Smart Manufacturing Leadership Coalition, 2013.

[15] Ray Kurzwail lecture "Biologically Inspired Models of Intelligence", June, 25, 2014. to the GOOGLE develpment engineering group

doi: n/a

Effect of Screw Configuration, Moisture Content and Particle Size of Corn Grits on Properties of Extrudates*

Antun Jozinović[1,], Drago Šubarić[1,**], Đurđica Ačkar[1,**], Jurislav Babić[1], Mirela Planinić[1], Mariana Pavoković[1], Marijana Blažić[2]**

[1]Josip Juraj Strossmayer University of Osijek, Faculty of Food Technology Osijek, Franje Kuhača 20, HR-31000 Osijek, Croatia
[2]Karlovac University of Applied Sciences, Ivana Meštrovica 10, HR-47000 Karlovac, Croatia

Extrusion is a modern procedure for processing different types of raw materials and production of wide range of food products, where the corn grits are often used as main raw materials. Therefore the aim of this study was to determine the effect of screw configuration (4:1 and 1:1), moisture content (15% and 20%) and particle size of corn grits (>500 μm and <500 μm) on properties of extrudates. Samples were extruded in the laboratory single screw extruder Brabender 19/20 DN, at temperature profile 135/170/170 °C, using die with 4 mm diameter. Physical and rheological properties, digestibility and starch damage of the obtained extrudates were determined, and results were compared with control samples of non-extruded corn grits. Lower moisture content and usage of screw with compression ratio 4:1 increased expansion ratio and fracturability, but decreased bulk density and hardness of extrudates, regardless of granularity. After extrusion process water absorption index increased, but peak, hot and cold viscosity of all samples decreased, with more pronounced effect in grits extruded with lower moisture content and with screw 4:1. Extrusion caused a reduction of the resistant starch content and increase starch damage o f all samples.

Key words:

extrusion, corn grits, particle size, moisture content, screw configuration

* The article was originally published in the „Croatian Journal of Food Science and Technology“ 2, vol.4, 2012 (str. 95-101) by the publisher Josip Juraj Strossmayer University of Osijek, Faculty of Food Technology, Osijek, Osijek, Croatia. The article has been approved for scholarly and non-commercial use.

?**ajozinovic@ptfos.hr, dsubaric@ptfos.hr, dackar@ptfos.hr

1. Introduction

Extrusion cooking of cereals is a very important process in food industry, since it regards a wide range of products such as snack-foods, baby-foods, breakfast cereals, noodle, pasta and cereals based blends (Semaska et al., 2010). Extruders minimize the operating costs and higher productivity than other cooking process, combining energy efficiency and versatility (Ficarella et al., 2004). Corn meal is a major ingredient for extruded foods, such as ready-to-eat breakfast cereals and snacks (Gujral et al., 2001). The effect of various process variables (moisture content, particle size, screw configuration) on extrusion behavior of corn grits and other flours have been extensively studied (Onwulata and Konstace, 2006; Delgado-Licon et al., 2009; Garber et al., 1997; Curic et al., 2009; Hood-Niefer and Tyler, 2010; Qi and Onwulata, 2011; Altan et al., 2009; Chuang and Yeh, 2004; Desrumaux et al., 1998; Carvalho et al., 2010). Rayas-Duarte et al. (1998) reported that samples prepared and extruded with the lowest moisture content have the largest water absorption and expansion index. Altan et al. (2009) investigated effect of screw configuration on properties of barley extrudates and reported that severe screw configuration produced more expanded product with low bulk density than that of medium screw configuration. Desrumaux et al. (1998) reported that increase in particle size for a given biochemical composition gave extrudates that were harder, with a modified expansion and lower cell density. Onwulata and Konstace (2006) reported that smaller particle size fractions exhibited increased solubility and significantly higher viscosity. Kebede et al. (2010) reported that increased barrel temperature, reduced feed moisture content and a higher screw speed showed a significantly ($P<0.01$) higher radial expansion, reduced bulk density and less compression resistance of teff flour extrudates. Carvalho et al. (2010) reported that extrudates produced with corn meal of higher particle sizes expanded more than extrudates produced with smaller particle sizes, and that increasing corn meal particle size decreased water absorption index (WAI) values.

The aim of this research was to determine the effect of screw configuration (4:1 and 1:1), moisture content (15% and 20%) and particle size of corn grits (>500 μm and <500 μm) on properties of extrudates – physical and rheological properties, digestibility and starch damage.

2. Materials and Methods

Corn grits used in this study, with different particle size (>500 μm and <500 μm), were obtained from the mill „Đakovo“ of the “Žito” Company Ltd. Osijek.

Moisture content of corn grits was set to 15% and 20%, and then samples were extruded in the laboratory single screw extruder Do-Coder, Brabender 19/20 DN, GmbH, Duisburg, Germany. *Extrusion parameters* were as follows: screw: 4:1 and 1:1; die: 4 mm; temperature profile: 135/170/170 °C. Obtained extrudates were air-dried.

Extrudate diameter and *expansion ratio (ER)* were measured according to Brnčić et al. (2008), where expansion ratio was calculated as follows (Eq. 1):

$$ER = \frac{\text{extrudate diameter (mm)}}{\text{die diameter (mm)}} \tag{1}$$

Bulk density (BD) of extrudates was measured according to Pan et al. (1998) and calculated according to Eq. 2:

$$BD = \frac{\text{extrudate mass (g)}}{\text{extrudate volume (cm}^3\text{)}} \tag{2}$$

Texture analysis was performed on texturometer TA.XT2 Plus, Stable Microsystem using method „Measurement of the hardness and fracturability of pretzel sticks“ with following settings: Pre-Test Speed: 1.0 mm/s; Test Speed: 1.0 mm/s; Post-Test Speed: 10.0 mm/s; Distance: 3 mm; Trigger Type: Auto – 5 g.

Water absorption index (WAI) was determined according to Sosulski (1962).

Pasting properties of extrudates (10% d. m., 100 g total weight) were measured using a Micro Visco-Analyser (Model 803202, Brabender Gmbh & Co KG, Duisburg, Germany). The flour suspensions were heated at 7.5 ºC/min from 32 to 92 ºC, held at 92 ºC for 10 min, cooled at 7.5 ºC/min to 50 ºC, and held at 50 ºC for 1 min.

Starch damage was determined according to AACC Method 76-31, and *resistant starch content* according to AOAC 2002.02 method.

Experimental data were analyzed by analysis of variance (ANOVA) and Fisher's least significant difference (LSD) with significance defined at $P<0.05$. All statistical analyses were carried out using software program STATISTICA 10.0 (StatSoft, Inc, USA).

3. Results and Discussion

Effect of screw configuration, moisture content and particle size of corn grits on properties of extrudates is shown in Fig. 1. Extrudates prepared at lower moisture content and with screw configuration 4:1 were more expanded, regardless of corn grits granularity. These results are in accordance with other researches (Kebede et al., 2010; Onwulata and Konstance, 2006; Altan et al., 2009). The cross-sectional expansion index (SEI) was increased with a decrease in water injection rate and/or an increase in barrel temperature (Ryu and Ng, 2001). Extrudates produced with corn meal of higher particle sizes expanded more than extrudates produced with smaller particle sizes (Carvalho et al., 2010).

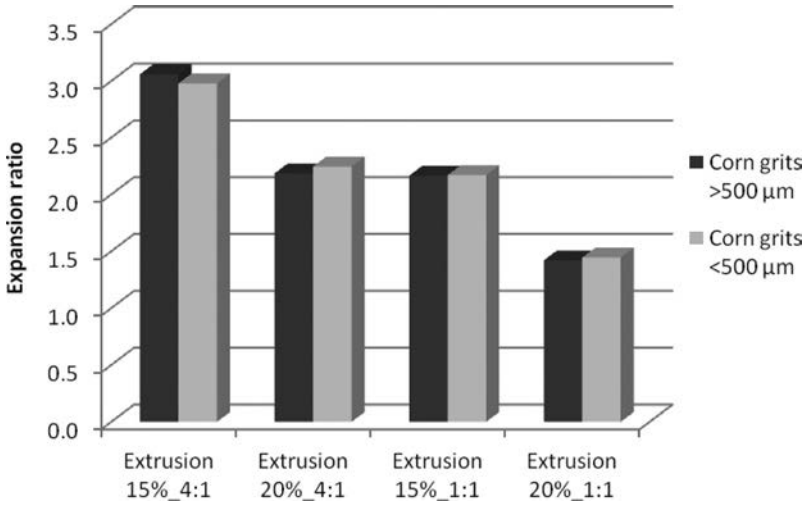


Fig. 1 – Expansion ratio of corn grits extrudates

Bulk density (BD) of extrudates was higher when samples were extruded at higher moisture content and with usage screw configuration 1:1 (Fig. 2). The results obtained by measuring the bulk density are in accordance with the results of measurements of the expansion ratio i.e. extrudates with lower values of the diameter and expansion ratio had higher bulk density. The lowest BD value was obtained when rice flour was extruded at lower moisture contents and higher temperatures, whereas the highest value was obtained at higher moisture contents and lower temperatures (Hagenimana et al., 2006). Other researches also show the same trend, where is concluded that an increase of moisture content increased bulk density of extrudates (Ding et al., 2005, 2006; Thymi et al., 2005; Garber et al., 1997), which confirms the results obtained in this investigation.

Texture properties of extrudates are highly influenced by expansion degree (Anton et al., 2009). This is in accordance with this research, since extrudates with lower

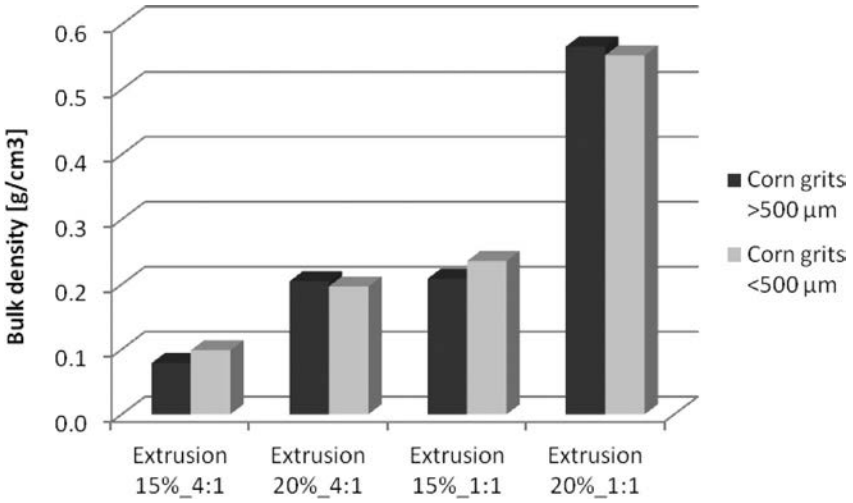


Fig. 2 – Bulk density of corn grits extrudates

expansion ratio had higher hardness and lower fracturability (Fig. 3; Fig. 4). Extrudates extruded at higher moisture content and with usage screw 1:1 had higher hardness and lower fracturability, regardless of corn grits granularity. Many studies have shown that the greatest impact on the texture of extrudates has moisture content (Brnčić et al., 2006; Petrova et al., 2010), but also other parameters (temperature, screw configuration, screw speed...) are very important and were subjects of many papers (Lazou and Krokida, 2010; Mendonca et al., 2000; Saeleaw et al., 2012, Wu et al., 2007).

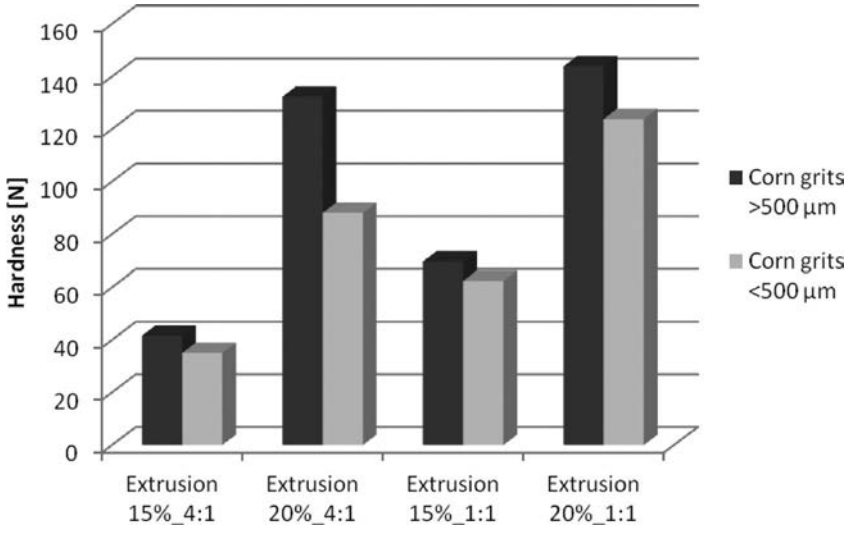


Fig. 3 – Hardness of corn grits extrudates

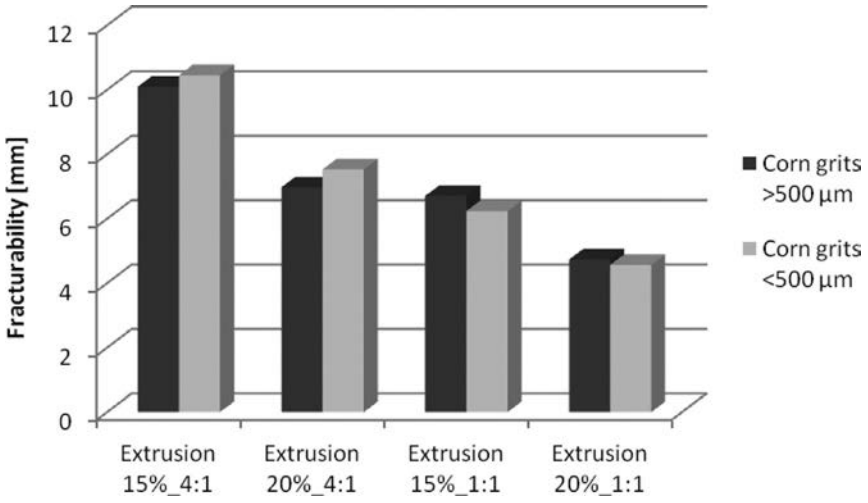


Fig. 4 – Fracturability of corn grits extrudates

Water absorption index (WAI) values of all extruded samples were significantly higher compared to non-extruded corn grits, and the greater values were in more expanded extrudates (Fig. 5). Ding et al. (2005; 2006) concluded that the increase in moisture content results in extrudates with lower water absorption index values. Increasing corn meal particle size decreased WAI values (Carvalho et al., 2010). One more study, which is in accordance with the results obtained in this work, concluded that extrusion causes an increase in water absorption index (Larrea et al., 2005).

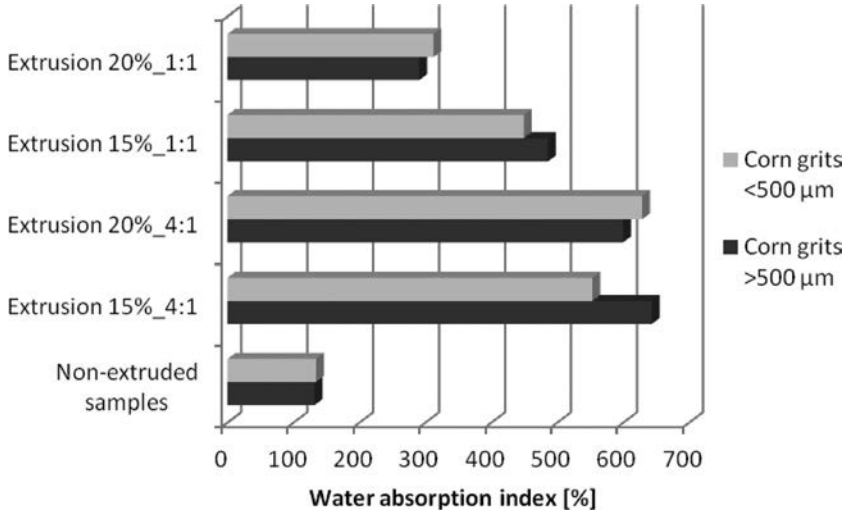


Fig. 5 – Water absorption index (WAI) of non-extruded and extruded corn grits

In Table 1 pasting properties of non-extruded and extruded corn grits are shown. Extrusion resulted in decrease of peak, hot and cold viscosity of all samples. Decrease of peak viscosity is correlated to effect of extrusion on higher degradation and gelatinisation of starch (Hagenimana et al., 2006). Peak viscosity was lower at higher moisture content in extrusion using screw 4:1, but on the other hand, in extrusion using screw 1:1 peak viscosity was lower at lower moisture content. Gutkoski and El-Dash (1999) reported that initial viscosity of the paste increased with the increase of raw material moisture and extrusion temperature, while the maximum viscosity (at a constant temperature) decreased with the increase of temperature. Peak viscosity of corn grits with higher particle size was lower in non-extruded samples, but higher in all extruded samples. After heating and mixing on 92 °C viscosity of all samples decreased, and then after cooling on 50 °C viscosity of all samples marked increased, as a result of starch retrogradation during cooling. Viscosity values at 50 °C of non-extruded samples were higher in regards to all extruded samples. This is in accordance to other researches (Hagenimana et al., 2006; McPherson et al., 2000). Extrusion process partially damages starch granules, thus obtained gels of extrudated products have lower viscosity than the initial grits (Dokić et al., 2009).

Table 1 – Pasting properties of non-extruded and extruded corn grits

Sample	Peak viscosity [BU]	Viscosity at 92 °C [BU]	After mixing at 92 °C [BU]	Viscosity at 50 °C [BU]
Non-extruded samples				
Corn grits >500 µm	484.5 ± 18.5^a	112 ± 4.0^a	477.5 ± 21.5^a	809 ± 21.0^a
Corn grits <500 µm	502 ± 1.0^a	342.5 ± 1.5^b	475.5 ± 1.5^a	877 ± 2.0^a
Extrusion 15%_4:1				
Corn grits >500 µm	253 ± 36.0^a	154.5 ± 6.5^a	153 ± 9.0^a	242 ± 6.0^a
Corn grits <500 µm	230.5 ± 9.5^a	141 ± 3.0^a	137 ± 1.0^a	209 ± 4.0^a
Extrusion 20%_4:1				
Corn grits >500 µm	212 ± 18.0^a	191 ± 20.0^a	185 ± 3.0^a	293.5 ± 6.5^a
Corn grits <500 µm	181.5 ± 9.5^a	155 ± 2.0^a	165 ± 4.0^a	279 ± 4.0^a
Extrusion 15%_1:1				
Corn grits >500 µm	130 ± 5.0^a	107.5 ± 3.5^b	116.5 ± 7.5^a	172.5 ± 3.5^b
Corn grits <500 µm	128 ± 2.0^a	89 ± 0.0^a	91 ± 1.0^a	145 ± 2.0^a
Extrusion 20%_1:1				
Corn grits >500 µm	255.5 ± 1.5^a	213.5 ± 0.5^a	254.5 ± 1.5^a	474 ± 3.0^a
Corn grits <500 µm	250 ± 1.0^a	233 ± 2.0^a	246 ± 0.0^a	444 ± 1.0^a

Starch damage significantly increased and resistant starch (RS) content decreased after extrusion of all samples (Fig. 6; Fig. 7). This is in accordance to conclusion of Mendoza and Bresaani (1987) and Hagenimana et al. (2006). RS content was higher, before and after extrusion, in corn grits with higher particle size.

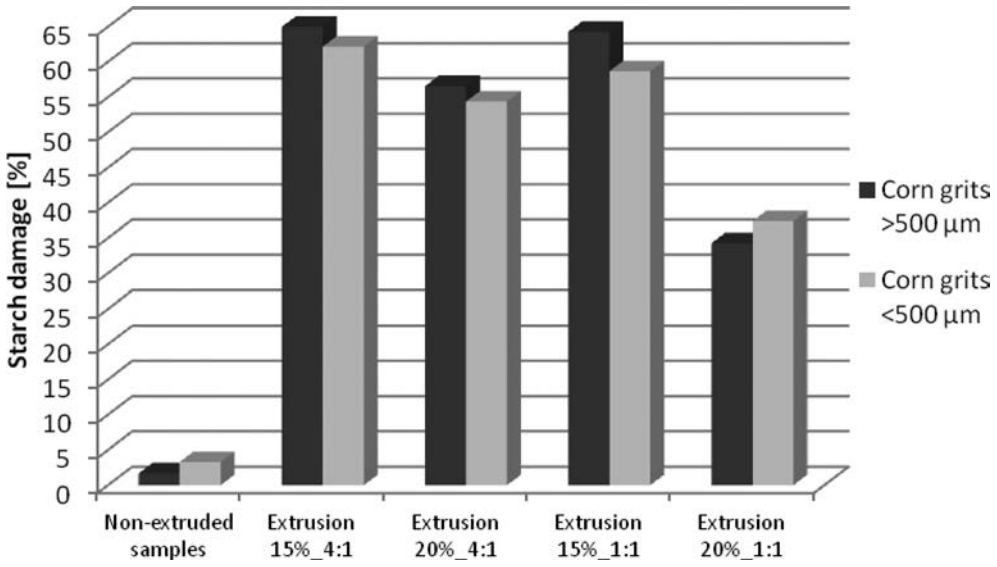


Fig. 6 – Starch damage of non-extruded and extruded corn grits

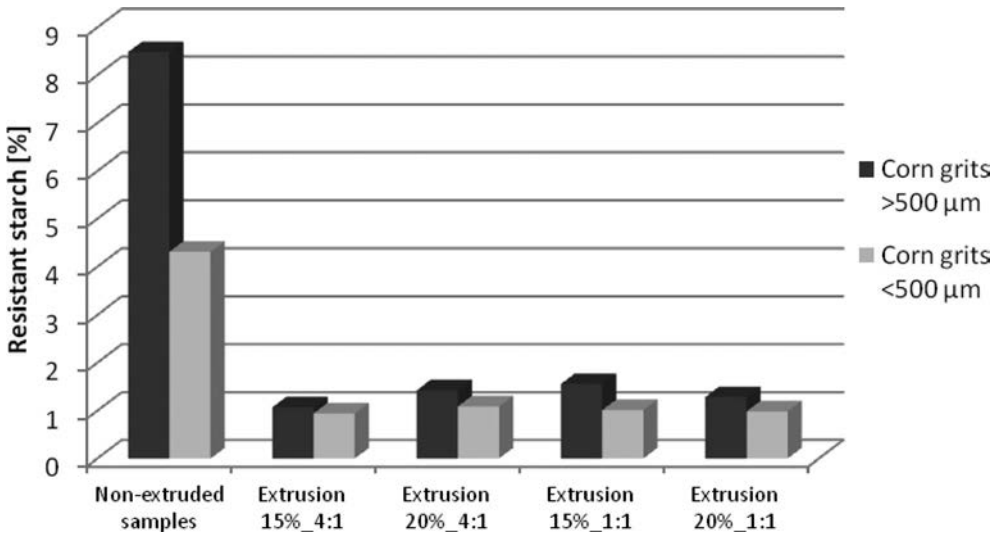


Fig. 7 – Resistant starch content of non-extruded and extruded corn grits

4. Conclusions

This research was conducted to investigate the parameters of extrusion for preparation corn grits „snack“ products with desirable physical properties.

Extrudates prepared at lower moisture content and with screw configuration 4:1 were more expanded, regardless of corn grits granularity. The results for bulk density and texture properties are in accordance with the results of measurements of the expansion ratio i.e. extrudates with lower values of the diameter and expansion ratio had higher bulk density and hardness and lower fracturability. Water absorption index values of all extruded samples were significantly higher compared to non-extruded corn grits. Extrusion resulted in decrease of peak, hot and cold viscosity of all samples. Starch damage significantly increased and resistant starch (RS) content decreased after extrusion.

The results obtained in this research indicate that screw with compression ratio 4:1 and corn grits with higher particle size and with lower moisture content results in products with better physical properties in regardless to products obtained with screw 1:1 and with corn grits with smaller particle size and with higher moisture content.

Acknowledgements

Results shown have outcome from scientific project “Development of new modified starches and their application in food industry” supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

References

[1] AACC 76-31, *Starch damage.* Official Methods of Analysis, of A.A.C.C. American Association of Cereal Chemists (AACC Method 76-31). The Association, St. Paul, MN.

[2] Altan, A., McCarthy, K.L., Maskan, M. (2009): Effect of screw configuration and raw material on some properties of barley extrudates. *Journal of Food Engineering* 92, 377–382.

[3] Anton, A.A., Fulcher, R.G., Arntfield, S.D. (2009): Physical and nutritional impact of fortification of corn-starch based extruded snacks with common bean (*Phaseolus vulgaris* L.) flour: effects of bean addition and extrusion cooking. *Food Chem.* 113, 989-996.

[4] AOAC 2002.02, *Resistant starch in starch and plant materials*. Official methods of analysis of the AOAC international (18th ed.). Gaithersburg, Maryland: AOAC International.

[5] Brnčić, M., Ježek, D., Rimac Brnčić, S., Bosiljkov, T., Tripalo, B. (2008): Utjecaj dodatka koncentrata proteina sirutke na teksturalna svojstva izravno ekspandiranog kukuruznog ekstrudata. *Mljekarstvo* 58 (2),131-149.

[6] Brnčić, M., Tripalo, B., Ježek, D., Semenski, D., Drvar, N., Ukrainczyk, M. (2006): Effect of twin-screw extrusion parameters on mechanical hardness of direct-expanded extrudates. *Sadhana* 31 (5), 527-536.

[7] Carvalho, C.W.P., Takeiti, C.Y., Onwulata, C.I., Pordesimo, L.O. (2010): Relative effect of particle size on the physical properties of corn meal extrudates: Effect of particle size on the extrusion of corn meal. *Journal of Food Engineering* 98, 103–109.

[8] Chuang, G.C.C., Yeh, A. (2004): Effect of screw profile on residence time distribution and starch gelatinization of rice flour during single screw extrusion cooking. *Journal of Food Engineering* 63, 21–31.

[9] Curic, D., Novotni, D., Bauman, I., Kricka, T., Dugum, J. (2009): Optimization of extrusion cooking of cornmeal as raw material for bakery products. *Journal of Food Engineering* 32, 294-317.

[10] Delgado-Licon, E., Martinez Ayala, A.L., Rocha-Guzman, N.E., Gallegos-Infanate, J.A., Atienzo-Lazos, M., Drzewicki, J., Martínez-Sánchez, C.E., Gorinstein, S. (2009): Influence of extrusion on the bioactive compounds and the antioxidant capacity of the bean/corn mixtures. *International Journal of Food Sciences and Nutrition* 60 (6), 522-532.

[11] Desrumaux, A., Bouvier, J.M., Burri, J. (1998): Corn Grits Particle Size and Distribution Effects on the Characteristics of Expanded Extrudates. *Journal of Food Science* 63 (5).

[12] Ding, Q.B., Ainsworth, P., Plunkett, A., Tucker, G., Marson, H. (2006): The effect of extrusion conditions on the functional and physical properties of wheat-based expanded snacks. *Journal of Food Engineering* 73, 142–148.

[13] Ding, Q.B., Ainsworth, P., Plunkett, A., Tucker, G., Marson, H. (2005): The effect of extrusion conditions on the physicochemical properties and sensory characteristics of rice-based expanded snacks. *Journal of Food Engineering* 66, 283-289.

[14] Dokić, LJ.P., Bodroža-Solarov, M.I., Hadnađev, M.S., Nikolić, I.R. (2009): Properties of extruded snacks supplemented with amaranth grain grits. *Biblid* 40, 17-24.

[15] Ficarella, A., Milanese, M., Laforgia, D. (2004): Numerical study of extrusion process in cereals production: Part I. *Fluid-dynamic analysis of extrusion system* 73, 103-111.

[16] Garber, B.W., Hsieh, F., Huff, H.E. (1997): Influence of Particle Size on the Twin-Screw Extrusion of Corn Meal. *Cereal Chem.* 74 (5), 656–661.

[17] Gujral, H. S., Singh, N., Singh, B. (2001): Extrusion behaviour of grits from flint and sweet corn. *Food Chemistry* 74, 303–308.

[18] Gutkoski, L.C., El-Dash, A.A. (1999): Effect of extrusion process variables on physical and chemical properties of extruded oat products. *Plant Foods for Human Nutrition* 54, 315–325.

[19] Hagenimana, A., Ding, X., Fang, T. (2006): Evaluation of rice flour modified by extrusion cooking. *Journal of Cereal Science* 43 (1), 38-46.

[20] Hood-Niefer, S.D., Tyler, R.T. (2010): Effect of protein, moisture content and barrel temperature on the physicochemical characteristics of pea flour extrudates. *Food Research International* 43, 659–663.

[21] Kebede, L., Worku, S., Bultosa, G., Yetneberek, S. (2010): Effect of extrusion operating conditions on the physical and sensory properties of tef (*Eragrostis tef* [Zucc.] Trotter) flour extrudates. *EJAST*, 1 (1) 27-38.

[22] Larreaa, M.A., Changb, Y.K., Martinez-Bustos, F. (2005): Some functional properties of extruded orange pulp and its effect on the quality of cookies. *LWT* 38, 213–220.

[23] Lazou, A., Krokida, M. (2010): Structural and textural characterization of corn–lentil extruded snacks. *Journal of Food Engineering* 100, 392–408.

[24] McPherson, A.E., Bailey, T.B., Jane, J. (2000): Extrusion of cross-linked hydroxypropylated corn starches I. Pasting properties. *Cereal Chem.* 77 (3), 320-325.

[25] Mendonca, S., Grossmann, M.V.E., Verhé, R. (2000): Corn Bran as a Fibre Source in Expanded Snacks. *Lebensm.-Wiss. u.-Technol.* 33, 2-8.

[26] Mendoza, C.M., Bressani, R. (1987): Nutritional and functional characteristics of extrusion-cooked amaranth flour. *Cereal Chem* 64 (4), 218-222.

[27] Onwulata, C.I., Konstance, R.P. (2006): Extruded corn meal and whey protein concentrate: Effect of particle size. *Journal of Food Processing and Preservation* 30, 475–487.

[28] Pan, Z., Zhang, S., Jane, J. (1998): Effects of extrusion variables and chemicals on the properties of starch-based binders and processing conditions. *Cereal Chemistry* 75, 541-546.

[29] Petrova, T., Ruskova, M., Tzonev, P., Zsivanovits, G., Penov, N. (2010): Effect of Extrusion Variables on the Hardness of Lentil Semolina Extrudates. In *7th International Conference of the Balkan Physical Union*, *AIP Conference Proceedings* 1203, 1031-1036.

[30] Qi, P.X., Onwulata, C.I. (2011): Physical properties, molecular structures, and protein quality of texturized whey protein isolate: Effect of extrusion moisture content. *Journal of Dairy Science* 94 (5), 2231- 2244.

[31] Rayas-Duarte, P., Majewska, K., Doetkott, C. (1998): Effect of Extrusion Process Parameters on the Quality of Buckwheat Flour Mixes. *Chereal Chemistry* 75 (3).

[32] Ryu, G.H., Ng, P.K.W. (2001): Effects of Selected Process Parameters on Expansion and Mechanical Properties of Wheat Flour and Whole Cornmeal Extrudates. *Starch/Stärke* 53, 147–154.

[33] Saeleaw, M., Dürrschmid, K., Schleining, G. (2012): The effect of extrusion conditions on mechanical-sound and sensory evaluation of rye expanded snack. *Journal of Food Engineering* 110, 532–540.

[34] Semaska, C., Kong, X., Hua, Y. (2010): Optimization of Extrusion on Blend Flour Composed of Corn, Millet and Soybean. *Pakistan Journal of Nutrition* 9 (3), 291-297.

[35] Sosulski, F.W. (1962): The centrifuge method for determining flour absorption in hard red spring wheat. *Cereal Chemistry* 39, 344-350.

[36] Thymi, S., Krokida, M.K., Pappa, A., Maroulis, Z.B. (2005): Structural properties of extruded corn starch. *Journal of Food Engineering* 68, 519–526.

[37] Wu, W., Huff, H.E., Hsieh, F. (2007): Processing and properties of extruded flaxseed-corn puff. *Journal of Food Processing and Preservation* 31, 211–226.

doi: n/a

Food Industry By-Products as Raw Materials in Functional Food Production

Antun Jozinović[1,], Drago Šubarić[1], Đurđica Ačkar[1], Borislav Miličević[1], Jurislav Babić[1,**], Midhat Jašić[2], Kristina Valek Lendić[3]**

[1]Josip Juraj Strossmayer University of Osijek, Faculty of Food Technology Osijek, Franje Kuhača 20, HR-31000 Osijek, Croatia
[2]University of Tuzla, Faculty of Technology, Univerzitetska 8, 75 000 Tuzla, Bosnia and Herzegovina
[3]Institute of Public Health of the Osijek-Baranja County, Franje Krežme 1, 31000 Osijek, Croatia

Western civilization problems nowadays are overweight, obesity, diabetes, cardiovascular diseases, cancer and different disorders closely linked to unbalanced diet.

Since it is extremely difficult to influence nutritional preferences of consumers, food industry is now increasingly developing new products, such as bread, pasta, snack products and other highly consumed products by all groups of consumers enriched with ingredients that are lacking in every day nutrition (fiber, polyphenols, antioxidants, vitamins, β-glucan…) and functional products which have scientifically proven beneficial effect on human health.

Food industry by-products, such as apple pomace, by-products from sugar industry and brewers spent grains are rich source of polyphenols, fiber and β-glucan. Grape pomace is rich in polyphenols, tomato pomace in lycopene and carrot pomace in β-carotene. These are just some examples of by-products with great potential of application in enriched and functional food production. In addition to natural substances which are produced in this manner, problem of large quantities of waste disposal is also resolved.

Keywords:

food industry by-products, functional food, fiber, β-glucan, antioxidants

[*] The article was originally published in the journal „Food in health and disease“ 1, vol.3., 2014 (22-30) by the publisher University of Tuzla, Faculty of Pharmacy, Tuzla, BH and co-publisher University of Osijek, Faculty of Food Technology, Osijek, Croatia. The article has been approved for scholarly and non-commercial use.
[**]ajozinovic@ptfos.hr, jbabic@ptfos.hr

1. Introduction

The modern problems of Western civilization are overweight, obesity, diabetes, cardiovascular diseases and various disorders that are closely related to improper diet. As it is difficult to affect eating habits of consumers, today's food industry develops new products consumed by the wide population, enriched with ingredients that are poorly represented in the daily diet (fiber, antioxidants, polyphenols, vitamins, β-glucan, minerals,...) and functional products, which are scientifically proven to have a beneficial effect on health. Modern trend is a demand and production of food products with the specific taste and health benefits. All of these requirements of consumers pose a major challenge for food technologists and all those involved in the food production chain.

By-products of plant food processing represent a major disposal problem for the industry concerned, but they are also promising sources of compounds which may be used because of their favourable technological or nutritional properties (Schieber et al., 2001b). At present up to one third of fruit and vegetables in the form of peels, pips, kernels and skins can be discarded during preparation and processing, therefore creating a 'waste', while also decreasing the maximum nutritional potential of the fruit or vegetable (O'Shea et al., 2012).

This review illustrates nutritional value of some food industry by-products and their application in production of various type of new products.

Table 1 – The content of various compounds in the fruit and vegetable by-products

Compound	Content [% w/w, db]	Source	References
Pectin	13 – 39	Apple pomace	Renard et al. (1996)
	15 – 30	Sugar beet pulp	Yapo et al. (2007)
Total dietary fibre • Insoluble • Soluble	51.1 36.5 14.6	Apple pomace	Sudha et al. (2007)
	57 47.6 9.41	Orange peel	Chau and Huang (2003)
	63.6 50.1 13.5	Carrot pomace	Chau et al. (2004)
Protein	27.5	Kernels of peach	Rahma (1988)
	20 – 25	Bitter apricot seeds	Tunçel et al. (1998)
	16.1	Cauliflower	Stojceska et al. (2008a)
	20	Brewer's spent grain	Mussatto et al. (2006)

2. Fruit by-products

Because of its high quantity in fruit processing industry and their nutritive value (dietary fibers, polyphenols, pectins,...) in this chapter are presented some of the most investigated fruit industry by-products.

2.1 Apple

The major product from apple processing is apple juice. The entire fruit is usually pressed in a cold press to extract the juice from the fruit. This can result in much waste, which is termed apple pomace (O'Shea et al., 2012). Apple pomace, inexpensive and primary by-product of apple juice and cider production is used as a source of pectin (Hwang et al., 1998), as animal feed (Sandhu and Joshi, 1997), as dietary fibres (Leontowicz et al., 2001) or as a source of phenolic compounds (Schieber et al., 2004).

Since apple pomace is rich in pectins, between 13 and 39 % of pectins (Renard et al., 1996), Royer et al. (2006) showed that it is possible to obtain jellies with apple pomace without incorporating gel additive. Production of pectin is considered the most reasonable way of utilizing apple pomace both from an economical and from an ecological point of view (Endreβ, 2000, Fox et al., 1991). In comparison to citrus pectins, apple pectins are characterized by superior gelling properties. However, the slightly brown hue of apple pectins caused by enzymatic browning may lead to limitations with respect to their use in very light-coloured foods (Schieber et al., 2001b).

Gorinstein et al. (2001b) investigated the dietary fibre levels of a whole apple, its pulp and its peel. Interestingly, they found that the majority of the total fibre was located in the peel of the apple (0.91 % fresh weight [FW]). The percentage of insoluble (0.46 % FW) to soluble fibre (0.43 % FW) was found to be well balanced in terms of receiving a health benefit. Dried apple pomace is considered as a potential food ingredient, having dietary fibre content of about 36.8 %, and has been used in apple pie filling and in oatmeal cookies (Carson et al., 1994).

Apple pomace has been shown to be a good source of polyphenols which are predominantly localized in the peels and are extracted into the juice to a minor extent. Major compounds isolated and identified include catechins, hydroxycinnamates, phloretin glycosides, quercetin glycosides, procyanidins, chlorogenic and caffeic acid, and phloridzin (Foo and Lu, 1999, Lommen et al., 2000, Lu and Foo, 1997, 1998, Schieber et al., 2001a, Garcia et al., 2009, Schieber et al., 2003).

Masoodi et al. (2002) studied cake making from apple pomace wheat flour blends at 5, 10 and 15 %, so as to enrich the cake with fibre content. Sudha et al. (2007) also investigated the addition of apple pomace in wheat flour at 5, 10 and 15 % levels and studied rheological characteristics and cake making. These authors concluded that the cakes prepared with apple pomace had pleasant fruity flavour, and had higher dietary fiber and phenol contents.

Recently, apple pomace is tried to be incorporated into other products, such as "snacks", which are highly consumed products by all groups of consumers (Karkle et al., 2012).

2.2 Grape

Grape (*Vitis* sp., Vitaceae) is one of the world's largest fruit crop with more than 60 million tons produced annually. About 80 % of the total crop is used in wine making and pomace represents approximately 20 % of the weight of grapes processed. From these data it can be calculated that grape pomace amounts to more than 9 million tons per year (Schieber et al., 2001b). Viniculture is an important agricultural activity in a lot of countries in southern Europe like in Spain, Italy and France and produces huge amounts of grape marc. This by-product consists mainly of skins and in certain case of seeds and some stalks. After extraction in the distilleries of wide range of products (ethanol, grape seed oil, anthocyanins and tartrate), the remaining pomace is currently not upgraded but used for composting or discarded in open areas potentially causing environmental problems. Considering the growing demand for green materials and components, agricultural by-products like pomace have an obvious potential as a renewable starting material (Rondeau et al., 2013). It is also used in the production of citric acid, methanol, ethanol and xanthan gum as a result of fermentation. The nutritional and compositional characteristics of grape pomace are known to vary, depending on the grape cultivar, growth climates and processing conditions (Deng et al., 2011).

Grape pomace has been shown to be a rich source of dietary fibre; its components mainly comprise of cellulose, small proportions of pectins and hemi-celluloses (Kammerer et al., 2005, González-Centeno et al., 2010).

Furthermore, grape pomace has also been evaluated as a source of antioxidants because of its high contents of polyphenols (Negro et al., 2003). Anthocyanins, catechins, flavonol glycosides, phenolic acids and alcohols and stilbenes are the principal phenolic constituents of grape pomace (Schieber et al., 2001b). Ruberto et al. (2007) carried out a study on the polyphenol content of Sicilian red grape pomace. The authors found that anthocyanins, flavonols and the phenolic acid, gallic acid, were the main polyphenols present.

In recent years grape pomace was used for production of different types of products. Altan et al. (2009) investigated the functional properties and *in vitro* starch digestibility of barley-based extrudates from fruit and vegetable by-products (tomato and grape pomace), and concluded that increasing level of both tomato and grape pomace led to reduction in starch digestibility.

Graphical optimization studies resulted in 155-160 °C, 4.47-6.57 % pomace level and 150-187 rpm screw speed as optimum variables to produce acceptable extrudates and the results suggest that grape pomace can be extruded with barley flour into an acceptable snack food (Altan et al., 2008b).

2.3 Peach and apricot

Peaches and apricots contain significant quantities of phenolics and carotenoids, components with various health benefits (Campbell and Padilla-Zakour, 2013).

Peach has been widely used around the world in the form of peach slices in syrup or just eaten as a dessert. The remnants from peach processing usually include the kernel and the peel. Over the years, these remnants have been used for their pectin as a thickener in jams; nowadays they are used commercially as a general thickener in foods (O'Shea et al., 2012). Págan and Ibarz (1999) described the recovery of pectin from fresh peach pomace**.** It is concluded that peach pectin is highly methoxylated and has favourable gelling properties (Págan et al., 1999). Kurz et al. (2008) characterized the cell wall polysaccharides of peaches and concluded that the main polysaccharides found were in the form of pectin.

Kernels of peach fruits contain 54.5 % and 27.5 % oil and protein, respectively, but ash and total carbohydrates were quite low (Rahma, 1988). Because of this high content of oil, Sánchez-Vicente et al. (2009) used peach seed as raw material for supercritical fluid extraction of oil. Furthermore, peach seeds may be used for the production of persipan (Schieber et al., 2001b).

Apricot is one of the most delicious and commercially traded fruits in the world. The plant is rich in mono- and polysaccharides, polyphenols, fatty acids and sterol derivatives, carotenoids, cyanogenic glucosides, and volatile components due to its appealing smell (Erdogan-Orhan and Kartal, 2011).

More than 650 metric tonnes of bitter apricot seeds are produced in Turkey per year as a by-product from the fruit canning industry (Tunçel, 1995). They are used as a substitute for bitter almonds to produce persipan for the bakery industry. The oil (53 % in the seed) is used, in e.g. cosmetics, as a cheaper substitute for bitter almond oil. The seeds can also be of interest as a food or feed ingredient because of

their high crude protein content (20-25 % w/w, dry weight basis). The main problem is that bitter seeds contain approximately 50-l50 μmol/g (dry weight basis) of potentially toxic cyanogenic glycosides, mainly amygdalin and prunasin (Tunçel et al., 1998). Because of that, before using, seeds must be debittered by hydrolysis of amygdalin (Schieber et al., 2001b), and there are various researches about this (Tunçel et al., 1990, 1995, 1998, Nout et al., 1995).

2.4 Lemon and orange

Approximately 50 % of the original whole fruit mass, after citrus processing for juice, consist of the peel, membranes and seeds. Citrus residues consist mainly of insoluble fiber (celluloses) and a small proportion of soluble fiber (hemicelluloses and pectin). For this reason, citrus residues could be considered as a potential high fiber ingredient that is used for food industry (García-Méndez et al., 2011). Residues of citrus juice production are a source of dried pulp and molasses, fiber-pectin, cold-pressed oils, essences, D-limonene, juice pulps and pulp wash, ethanol, seed oil, pectin, limonoids and flavonoids (Schieber et al., 2001b).

Comparison of some biochemical characteristics of different citrus fruits investigated Gorinstein et al. (2001a). These authors concluded that lemons possess the highest antioxidant potential among the studied citrus fruits and are preferable for dietary prevention of cardiovascular and other diseases. The peels of all citrus fruits are rich in dietary fibres and phenolic compounds and suitable for industrial processing.

García-Méndez et al., (2011) found that extrusion is a process that has the capability to transform insoluble fiber to soluble fiber in lemon residues. The highest content of soluble fiber was 50 %, when operating conditions were high in temperature (100 °C), low in moisture content (40 %) and low in screw speed (10 rpm).

85 % of oranges are processed into some form of orange juice, leaving behind tonnes of by-product after production. As a result of the functional and nutritional characteristics of orange peel, it may be considered to be a viable ingredient for a wide variety of products such as meat pastes, baked goods and yoghourt (O'Shea et al., 2012). Chau and Huang (2003) found that the orange peel contain 57 % DW total dietary fibre; of this 47.6 % DW was the insoluble fraction and 9.41 % DW was the soluble fraction.

Larrea et al. (2005) investigated the effects of some operational extrusion parameters on selected functional properties of orange pulp and its use in the preparation of biscuit-type cookies. They concluded that biscuits of good technological quality

and with a good level of acceptance were obtained by means of replacing up to 15 g/100 g of the wheat flour with extruded orange pulp.

3. Vegetable by-products

As a rich source of lycopene (tomato), β-carotene (carrot) and dietary fiber (cauliflower), and because of their high quantity in vegetable processing industry in this chapter are presented these three nutritive valuable vegetable industry by-products.

3.1 Tomato

Tomato (*Lycopersicon esculentum*) is one of the most popular vegetables and an integral part of human diet worldwide. Significant amounts are consumed in the form of processed products such as juice, paste, puree, ketchup, sauce and salsa (Altan et al., 2008a). During tomato processing a by-product, known as tomato pomace, is generated. This by-product represents, at most, 4 % of the fruit weight, and mainly consists of fibre; it can represent up to 50 % of the by-product on a dry weight basis (Del Valle et al., 2006). Furthermore, this by-product can still contain many nutrients and phytochemicals (O'Shea et al., 2012). The skin, important component of pomace, is source of lycopene. Lycopene is an excellent natural food color and also serves as a functional ingredient with important health benefits beyond basic nutrition (Kaur et al., 2005). It has been associated with various health benefit claims including immune system modulation, as a free radical scavenger and as having anticarcinogen properties (Dehghan-Shoar et al., 2010).

Nowadays, there are many researches about using tomato pomace as a novel ingredient in different types of food products. Dehghan-Shoar et al. (2010) investigated the addition of tomato derivatives to traditional starchy extruded snacks to improve their nutritional properties. These authors concluded that lycopene retention was higher in products containing tomato skin powder and significantly lower when wheat flour was used to make the snacks. Increases in the processing temperature improved the physicochemical characteristics of the snacks but had no significant effect on lycopene retention ($P > 0.05$) and texture of the product. Calvo et al. (2008) incorporated tomato powder (from tomato peel) into fermented sausages. Besides, tomato peel was successfully added to hamburgers to improve their nutritional content via the presence of lycopene (García et al., 2009). Tomato pomace can be extruded with barley flour into an acceptable and nutritional snack. Extrudates with 2 % and 10 % tomato pomace levels extruded at 160 °C and 200 rpm

had higher preference levels for parameters of color, texture, taste and overall acceptability (Altan et al., 2008a).

3.2 Carrot

The carrot (*Daucus carota*) is a root vegetable, usually orange, purple, red, white or yellow in color, with a crisp texture when fresh. It is a rich source of β-carotene and contains other vitamins, like thiamine, riboflavin, vitamin B-complex and minerals. Carrot pomace is a by-product obtained during carrot juice processing. The juice yield in carrots is only 60-70 %, and even up to 80 % of carotene may be lost with left over carrot pomace (Kumar et al., 2010). The total dietary fibre content of the carrot pomace was found to be 63.6 % DM, with 50.1 % DM being the insoluble fraction and 13.5 % DM the soluble fraction (Chau et al., 2004).

Because pomace received from carrots doesn't contain kernels and seeds, it can easily be added to a product without introducing negative functional or flavour issues while still retaining a lot of its phytochemicals (Chantaro et al., 2008). Various attempts were made at utilizing carrot pomace in food such as bread, cakes, dressing and for the production of functional drinks (Schieber et al., 2001b). Upadhyay et al. (2010) investigated the optimization of carrot pomace powder (CPP) incorporation on extruded product quality. The study demonstrated that an acceptable extruded product can be prepared by CPP incorporation, and optimum incorporation level of CPP was found to be 5 %. Kumar et al. (2010) found that carrot pomace could be incorporated into ready-to-eat expanded products up to the level of 8.25 %.

3.3 Cauliflower

Cauliflower has a very high waste index and is an excellent source of protein (16.1 %), cellulose (16 %) and hemicellulose (8 %). It is considered as a rich source of dietary fibre and it possesses both antioxidant and anticarcinogenic properties. Encouraging characteristics such as its pale colour, bland taste and high nutritional content make it an attractive novel ingredient (Stojceska et al., 2008a). Llorach et al. (2003) analysed the antioxidant capacity of cauliflower by-products and found that flavonoids and hydroxycinnamic acids were the main phenolics present. Similar to some of the fruit and vegetables, cauliflower by-products (such as the stem) have been shown to contain a significant amount of phytochemicals (O'Shea et al., 2012), and can be good novel ingredient for production of various food products. Stojceska et al., (2008a) used cauliflower by-products in production of cereal based ready-to-eat expanded snack and found that increasing the cauliflower to levels of 5-20 % increased dietary fibre in the finished product by over 100 %, increased protein content and water absorption index. Sensory test panel indicated that cauli-

flower could be incorporated into ready-to-eat expanded products up to the level of 10 %.

4. Sugar beet by-products

A third of the world production of sugar comes from sugar beet (*Beta vulgaris*). One ton of sugar beet (sucrose content 16 %) provides a dried weight of around 130 kg of sugar and 50 kg of a by-product, sugar beet pulp (SBP) (Rouilly et al., 2006). Molasses represents the runoff syrup from the final stage of crystallization. It mainly consists of fermentable carbohydrates (sucrose, glucose, fructose), and of nonsugar compounds which were not precipitated during juice purification. Molasses is used as feed and as a source of carbon in fermentation processes, e.g. for the production of alcohol, citric acid, L-lysine and L-glutamic acid (Schieber et al., 2001b).

Sugar beet pulp (SBP), a major by-product of the sugar refining industry, is a potential feedstock for biofuels. It contains 20-25 % cellulose, 25-36 % hemicellulose, 20-25 % pectin, 10-15 % protein, and 1-2 % lignin content on a dry weight basis (Zheng et al., 2013). Due to highly digestible fiber it is valued as an excellent food complement for animal feed and energy source. Raw pulp has been proposed as cultivation substrate, as well, for divalent cations complexation, as source of polyols for the production of urethanes and polyurethanes, as source of fiber in biodegradable composites or for paper manufacture (Rouilly et al., 2009). Addition of sugar beet fiber to semolina increased dietary fiber content but adversely affected colour and cooking loss of spaghetti (Özboy and Köksel, 2000). Owing to its high pectin content (15-30 %) on dry weight basis, and its availability in large quantities, sugar beet pulp (SBP) is another source, after apple pomace and citrus peels, for commercial pectin production (Yapo et al., 2007). Because of that, there are many researches about extraction of pectin from SBP (Li et al., 2012, Yapo et al., 2007, Lv et al., 2013, Ma et al., 2013). Pectins from SBP have poor gelling properties compared to citrus and apple pectins due to their high degree of methylation and low molecular weight and they are not extensively used in traditional applications in the food industry (Mata et al., 2009). In many parts of the world, utilization of SBP is an economically marginal part of beet sugar processing due to the low feed value and high drying cost. In certain areas, dehydrating and pelletizing SBP contribute 30-40 % of the overall energy cost of sugar beet processing. Therefore, the beet sugar industry seeks to add value to SBP via a process that does not require drying. In light of this, converting SBP into fuel ethanol through biological pathways, including hydrolysis and fermentation, is an attractive option (Zheng et al., 2012).

5. Brewer's spent grain

Brewer's spent grain (BSG) is the major by-product of the brewing industry. BSG is a lignocellulosic material containing about 17 % cellulose, 28 % non-cellulosic polysaccharides, chiefly arabinoxylans, and 28 % lignin. BSG is available in large quantities throughout the year, but its main application has been limited to animal feeding. Nevertheless, due to its high content of protein and fibre (around 20 and 70 % dry basis, respectively), it can also serve as an attractive adjunct in human nutrition (Mussatto et al., 2006). According to these authors, BSG is good for the manufacture of flakes, whole-wheat bread, biscuits and aperitif snacks, but it must be first converted to flour. Nevertheless, there are some limitations in the use of the flour as a protein additive or as a partial replacement for presently used flours, due to its colour and flavour. However, β-glucan from BSG has a significant positive impact on health and because of that BSG is excellent raw material for the production of functional products.

Recent researches show that BSG contains a significant content of polyphenols (Moreira et al., 2013, Meneses et al., 2013, McCarthy et al., 2012). Stojceska and Ainsworth (2008) added BSG in wheat flour in bread production. Increasing the level of dietary fibre increased dough development time, dough stability and crumb firmness but decreased the degree of softening and loaf volume. Ktenioudaki et al. (2013) investigated sensory properties and aromatic composition of baked snacks containing brewer's spent grain. They found that addition of BSG altered the odour profile of the snacks, however sensory results indicated that BSG-containing snacks at a level of 10 % were highly acceptable and highlighted the possibility of using BSG as a baking ingredient in the formulation of enhanced fibre baked snacks. Stojceska et al. (2008b) incorporated BSG into ready-to-eat expanded products and concluded that addition of BSG significantly increased protein content, phytic acid and bulk density. Furthermore, Ainsworth et al. (2007) found that addition of BSG in maize extrudates has no significant effect on the total antioxidant capacity (TAC) and total phenolic compounds (TPC) values, but increase phitic acid (PA), protein in vitro digestibility (PIVD) and resistant starch (RS) values.

6. Conclusions

The food processing industry produces large quantities of waste products. These by-products are sources of components of high nutritive value, and can be used as raw materials for other purposes. Furthermore, they are inexpensive and available in large quantities. This paper clearly demonstrates the high nutritional value that many by-products possess, and their application in production of various new products.

Acknowledgements

This work has been fully supported by Croatian Science Foundation under the project 1321.

References

[1] Ainsworth, P., Ibanoglu, S., Plunkett, A., Ibanoglu, E., Stojceska, V. (2007): Effect of brewers spent grain addition and screw speed on the selected physical and nutritional properties of an extruded snack. *Journal of Food Engineering,* 81, 702-709.

[2] Altan, A., McCarthy, K. L., Maskan, M. (2008a): Evaluation of snack foods from barley-tomato pomace blends by extrusion processing. *Journal of Food Engineering,* 84, 231-242.

[3] Altan, A., McCarthy, K. L., Maskan, M. (2008b): Twin-screw extrusion of barley-grape pomace blends: Extrudate characteristics and determination of optimum processing conditions. *Journal of Food Engineering*, 89(1), 24-32.

[4] Altan, A., McCarthy, K. L., Maskan, M. (2009): Effect of extrusion cooking on functional properties and in vitro starch digestibility of barley-based extrudates from fruit and vegetable by-products. *Journal Food Science*, 74(2), 77-86.

[5] Calvo, M. M., García, M. L., Selgas, M. D. (2008): Dry fermented sausages enriched with lycopene from tomato peel. *Meat Science,* 80(2), 167-172.

[6] Campbell, O. E., Padilla-Zakour, O. I. (2013): Phenolic and carotenoid composition of canned peaches (*Prunus persica*) and apricots (*Prunus armeniaca*) as affected by variety and peeling. *Food Research International*, 54, 448-455.

[7] Carson, K. J., Collins, J. L., Penfield, M. P. (1994): Unrefined, dried apple pomace as a potential food ingredient. *Journal of Food Science,* 59, 1213-1215.

[8] Chantaro, P., Devahastin, S., Chiewchan, N. (2008): Production of antioxidant high dietary fiber powder from carrot peels. *LWT- Food Science and Technology,* 41(10), 1987-1994.

[9] Chau, C. F., Chen, C. H., Lee, M. H. (2004): Comparison of the characteristics, functional properties, and in vitro hypoglycemic effects of various carrot insoluble fiber-rich fractions. *LWT-Food Science and Technology,* 37(2), 155-160.

[10] Chau, C. F., Huang, Y. L. (2003): Comparison of the chemical composition and physicochemical properties of different fibers prepared from the peel of *Citrus sinensis* L. Cv. Liucheng. *Journal of Agricultural and Food Chemistry,* 51(9), 2615-2618.

[11] Dehghan-Shoar, Z., Hardacre, A. K., Brennan, C. S. (2010): The physico-chemical characteristics of extruded snacks enriched with tomato lycopene. *Food Chemistry,* 123, 1117-1122

[12] Del Valle, M., Camara, M., Torija, M. E. (2006): Chemical characterization of tomato pomace. *Journal of the Science of Food and Agriculture,* 86, 1232-1236.

[13] Deng, Q., Penner, M. H., Zhao, Y. (2011): Chemical composition of dietary fiber and polyphenols of five different varieties of wine grape pomace skins. *Food Research International,* 44(9), 2711-2719.

[14] Endreb, H. U. (2000): Gehobene Qualität durch Produkt-Integrierten Umweltschutz–PIUS. *Flüssiges Obst,* 67, 460-463.

[15] Erdogan-Orhan, I., Kartal, M. (2011): Insights into research on phytochemistry and biological activities of *Prunus armeniaca* L. (apricot). *Food Research International,* 44, 1238-1243.

[16] Foo, L. Y., Lu, Y. (1999): Isolation and identification of procyanidins in apple pomace. *Food Chemistry,* 64, 511-518.

[17] Fox, G. F., Asmussen, R., Fischer, K., Endreb, H. U. (1991): Aufwand und Nutzen der Apfeltresterverwertung. *Flüssiges Obst,* 58, 492-499.

[18] García, M. L., Calvo, M., Selgas, M. D. (2009): Beef hamburgers enriched in lycopene using dry tomato peel as an ingredient. *Meat Science,* 83(1), 45-49.

[19] Garcia, Y. D., Valles, B. S., Lobo, A. P. (2009): Phenolic and antioxidant composition of by-products from the cider industry: Apple pomace. *Food Chemistry,* 117(4), 731-738.

[20] García-Méndez, S., Martínez-Flores, H. E., Morales-Sánchez, E. (2011): Effect of extrusion parameters on some properties of dietary fiber from lemon (*Citrus aurantifolia* Swingle) residues. *African Journal of Biotechnology,* 10(73), 16589-16593.

[21] González-Centeno, M. R., Rosselló, C., Simal, S., Garau, M. C., Lopez, F., Femenia, A. (2010): Physico-chemical properties of cell wall materials obtained from ten grape varieties and their byproducts: Grape pomaces and stems. *LWT- Food Science and Technology,* 43(10), 1580-1586.

[22] Gorinstein, S., Martin-Belloso, O., Park, Y. S., Haruenkit, R., Lojek, A., Ĉíž, M., Caspi, A., Libman, I., Trakhtenberg, S. (2001a): Comparison of some biochemical characteristics of different citrus fruits. *Food Chemistry,* 74(3), 309-315.

[23] Gorinstein, S., Zachwieja, Z., Folta, M., Barton, H., Piotrowicz, J., Zemser, M., et al. (2001b): Comparative contents of dietary fiber, total phenolics, and minerals in persimmons and apples. *Journal of Agricultural and Food Chemistry*, 49(2), 952-957.

[24] Hwang, J. K., Kim, C. J., Kim, C. T. (1998): Extrusion of apple pomace to facilitates pectin extraction. *Journal of Food Science,* 63, 841-844.

[25] Kammerer, D. R., Schieber, A., Carle, R. (2005): Characterization and recovery of phenolic compounds from grape pomace-A review. *Journal of Applied Botany and Food Quality,* 79, 189-196.

[26] Karkle, E. L., Alavi, S., Dogan, H. (2012): Cellular architecture and its relationship with mechanical properties in expanded extrudates containing apple pomace. *Food Research International,* 46 10-21.

[27] Kaur, D., Sogi, D. S., Gary, S. K., Bawa, A. S. (2005): Flotation-cumsedimentation system for skin and seed separation from tomato pomace. *Journal of Food Engineering,* 71, 341-344.

[28] Ktenioudaki, A., Crofton, E., Scannell, A. G. M., Hannon, J. A., Kilcawley, K. N., Gallagher, E. (2013): Sensory properties and aromatic composition of baked snacks containing brewer's spent grain. *Journal of Cereal Science,* 57(3), 384-390.

[29] Kumar, N., Sarkar, B. C., Sharma, H. K. (2010): Development and characterization of extruded product of carrot pomace, rice flour and pulse powder. *African Journal of Food Science,* 4(11), 703-717.

[30] Kurz, C., Carle, R., Schieber, A. (2008): Characterisation of cell wall polysaccharide profiles of apricots (*Prunus armeniaca* L.), peaches (*Prunus persica* L.), and pumpkins (*Cucurbita* sp.) for the evaluation of fruit product authenticity. *Food Chemistry,* 106, 421-430.

[31] Larrea, M. A., Chang, Y. K., Martinez-Bustos, F. (2005): Some functional properties of extruded orange pulp and its effect on the quality of cookies. *LWT- Food Science and Technology,* 38(3), 213-220.

[32] Leontowicz, M., Gorinstein, S., Bartnikowska, E., Leontowicz, H., Kulasek, G., Trakhtenberg, S. (2001): Sugar beet pulp and apple pomace dietary fibers improve lipide metabolism in rats fed cholesterol. *Food Chemistry,* 72, 73-78.

[33] Li, D., Jia, X., Wei, Z., Liu, Z. (2012): Box–Behnken experimental design for investigation of microwave-assisted extracted sugar beet pulp pectin. *Carbohydrate Polymers,* 88(1), 342-346.

[34] Llorach, R., Espín, J. C., Tomás-Barberán, F. A., Ferreres, F. (2003): Valorization of cauliflower (*Brassica oleracea* L. var. botrytis) by-products as a source of antioxidant phenolics. *Journal of Agricultural and Food Chemistry,* 51(8), 2181-2187.

[35] Lommen, A., Godejohann, M., Venema, D. P., Hollman, P. C. H., Spraul, M. (2000): Application of directly coupled HPLC-NMR-MS to the identification and confirmation of quercetin glycosides and phloretin glycosides in apple peel. *Analytical Chemistry*, 72, 1793-1797.

[36] Lu, Y., Foo, L. Y. (1997). Identification and quantification of major polyphenols in apple pomace. *Food Chemistry,* 59, 187-194.

[37] Lu, Y., Foo, L. Y. (1998). Constitution of some chemical components of apple seed. *Food Chemistry,* 61, 29-33.

[38] Lv, C., Yong Wang, Y., Wang, L., Li, D., Adhikari, B. (2013): Optimization of production yield and functional properties of pectin extracted from sugar beet pulp. *Carbohydrate Polymers,* 95(1), 233-240.

[39] Ma, S., Yu, S., Zheng, X., Wang, X., Bao, Q., Guo, X. (2013): Extraction, characterization and spontaneous emulsifying properties of pectin from sugar beet pulp. *Carbohydrate Polymers,* 98(1), 750-753.

[40] Masoodi, F. A., Bhavana, S., Chauhan, G. S. (2002): Use of apple pomace as a source of dietary fiber in cakes. *Plant Foods for Human Nutrition,* 57, 121-128.

[41] Mata, Y.N., Blázquez, M.L., Ballester, A., González, F., Muñoz, J.A. (2009): Sugar-beet pulp pectin gels as biosorbent for heavy metals: Preparation and determination of biosorption and desorption characteristics. *Chemical Engineering Journal,* 150(2-3), 289-301.

[42] McCarthy, A. L., O'Callaghan, Y. C., Connolly, A., Piggott, C. O., FitzGerald, R. J., O'Brien, N. M. (2012): Phenolic extracts of brewers' spent grain (BSG) as functional ingredients-Assessment of their DNA protective effect against oxidant-induced DNA single strand breaks in U937 cells. *Food Chemistry,* 134(2), 641-646.

[43] Meneses, N. G. T., Martins, S., Teixeira, J. A., Mussatto, S. I. (2013): Influence of extraction solvents on the recovery of antioxidant phenolic compounds from brewer's spent grains. *Separation and Purification Technology,* 108, 152-158.

[44] Moreira, M. M., Morais, S., Carvalho, D. O., Barros, A. A., Delerue-Matos, C., Guido, L. F. (2013): Brewer's spent grain from different types of malt: Evaluation of the antioxidant activity and identification of the major phenolic compounds. *Food Research International,* 54(1), 382-388.

[45] Mussatto, S.I., Dragone, G., Roberto, I.C. (2006): Brewers' spent grain: generation, characteristics and potential applications. *Journal of Cereal Science,* 43, 1-14.

[46] Negro, C., Tommasi, L., Miceli, A. (2003): Phenolic compounds and antioxidant activity from red grape marc extracts. *Bioresource Technology,* 87, 41-44.

[47] Nout, M. J. R., Tunçel, G., Brimer, L. (1995): Microbial degradation of amygdalin of bitter apricot seeds (*Prunus armeniaca*). International Journal of Food Microbiology, 24(3), 407-412.

[48] O'Shea, N., Arendt, E. K., Gallagher, E. (2012): Dietary fibre and phytochemical characteristics of fruit and vegetable by-products and their recent applications as novel ingredients in food products. *Innovative Food Science and Emerging Technologies*, 16, 1-10.

[49] Özboy, Ö., Köksel, H (2000): Effects of sugar beet fiber on spaghetti quality. *Zuckerindustrie,* 125, 248-250.

[50] Págan, J., Ibarz, A. (1999): Extraction and rheological properties of pectin from fresh peach pomace. *Journal of Food Engineering*, 39, 193-201.

[51] Págan, J., Ibarz, A., Llorca, M., Coll, L. (1999): Quality of industrial pectin extracted from peach pomace at different pH and temperatures. *Journal of the Science of Food and Agriculture,* 79, 1038-1042.

[52] Rahma, E. H. (1988): Chemical characterization of peach kernel oil and protein: Functional properties, *in vitro* digestibility and amino acids profile of the flour. *Food Chemistry,* 28(1), 31-43.

[53] Renard, C. M. G. C., Rohou, Y., Hubert, C., Della Valle, G., Thibault, J. F., Savina, J. P. (1996): Bleaching of apple pomace by hydrogen peroxide in alkaline conditions: optimisation and characterisation of the product. *Lebensmittel-Wissenchaft Und Technologie*, 30, 398-405.

[54] Rondeaua, P., Gambiera, F., Jolibert, F., Brossea, N. (2013): Compositions and chemical variability of grape pomaces from French vineyard. *Industrial Crops and Products*, 43, 251-254.

[55] Rouilly, A., Geneau-Sbartaï, C., Rigal, L. (2009): Thermo-mechanical processing of sugar beet pulp. III. Study of extruded films improvement with various plasticizers and cross-linkers. *Bioresource Technology,* 100, 3076-3081.

[56] Rouilly, A., Jorda, J., Rigal, L. (2006): Thermo-mechanical processing of sugar beet pulp. I. Twin-screw extrusion process. *Carbohydrate Polymers,* 66, 81-87.

[57] Royer, G., Madieta, E., Symoneaux, R., Jourjon, F. (2006): Preliminary study of the production of apple pomace and quince jelly. *LWT- Food Science and Technology,* 9(9), 1022-1025.

[58] Ruberto, G., Renda, A., Daquino, C., Amico, V., Spatafora, C., Tringali, C., et al. (2007): Polyphenol constituents and antioxidant activity of grape pomace extracts from five Sicilian red grape cultivars. *Food Chemistry,* 100(1), 203-210.

[59] Sánchez-Vicente, Y., Cabañas, A., Renuncio, J. A. R., Pando, C. (2009): Supercritical fluid extraction of peach (*Prunus persica*) seed oil using carbon dioxide and ethanol. *The Journal of Supercritical Fluids,* 49(2), 167-173.

[60] Sandhu, D. K. and Joshi, V. K. (1997): Solid state fermentation of apple pomace for concomitant production of ethanol and animal feed. *Journal of Scientific and Industrial Research*, 56, 86-90.

[61] Schieber, A., Hilt, P., Berardini, N., Carle, R. (2004): Apple pomace and mango peels as a source of pectin and phenolic compounds. In K. Waldron, C. Faulds, A. Smith (Eds.), *Proceeding Hrsg, Total Food*, Norwich, UK, pp. 145-156.

[62] Schieber, A., Hilt, P., Streker, P., Endress, H., Rentschler, C., Carle, R. (2003): A new process for the combined recovery of pectin and phenolic compounds from apple pomace. *Innovative Food Science & Emerging Technologies,* 4, 99-107.

[63] Schieber, A., Keller, P., Carle, R. (2001a): Determination of phenolic acids and flavonoids of apple and pear by high-performance liquid chromatography. *Journal of Chromatography A*, 910, 265-273.

[64] Schieber, A., Stintzing, F. C., Carle, R. (2001b): By-products of plant food processing as a source of functional compounds-Recent developments. *Trends in Food Science & Technology*, 12, 401-413.

[65] Stojceska, V., Ainsworth, P. (2008): The effect of different enzymes on the quality of high-fibre enriched brewer's spent grain breads. *Food Chemistry,* 110(4), 865-872.

[66] Stojceska, V., Ainsworth, P., Plunkett, A., İbanoğlu, E., İbanoğlu, Ş. (2008a): Cauliflower by-products as a new source of dietary fibre, antioxidants and proteins in cereal based ready-to-eat expanded snacks. *Journal of Food Engineering,* 87(4), 554-563.

[67] Stojceska, V., Ainsworth, P., Plunkett, A., Ibanoglu, S. (2008b): The recycling of brewer's processing by-product into ready-to-eat snacks using extrusion technology. *Journal of Cereal Science,* 47, 469-479.

[68] Sudha, M. L., Baskaran, V., Leelavathi, K. (2007): Apple pomace as a source of dietary fiber and polyphenols and its effect on the rheological characteristics and cake making. *Food Chemistry*, 104(2), 686-692.

[69] Tunçel, G. (1995): The effects of grinding, soaking and cooking on the degradation of amygdalin of bitter apricot seeds. *Food Chemistry,* 53(4), 447-451.

[70] Tunçel, G., Nout, M. J. R., Brimer, L. (1998): Degradation of cyanogenic glycosides of bitter apricot seeds (*Prunus armeniaca*) by endogenous and added enzymes as affected by heat treatments and particle size. *Food Chemistry,* 63(1), 65-69.

[71] Tun**ç**el, G., Nout, M. J. R., Brimer, L. (1995): The effects of grinding, soaking and cooking on the degradation of amygdalin of bitter apricot seeds. *Food Chemistry,* 53, 447-451.

[72] Tunçel, G., Nout, M. J. R., Brimer, L., Göktan, D. (1990): Toxicological, nutritional and microbiological evaluation of tempe fermentation with *Rhizopus oligosporus* of bitter and sweet apricot seeds. *International Journal of Food Microbiology,* 11(3-4), 337-344.

[73] Upadhyay, A., Sharma, H. K., Sarkar, B. C. (2010): Optimization of carrot pomace powder incorporation on extruded product quality by response surface methodology. *Journal of Food Quality,* 33(3), 350-369.

[74] Yapo, B. M., Robert, C., Etienne, I., Wathelet, B., Paquot, M. (2007): Effect of extraction conditions on the yield, purity and surface properties of sugar beet pulp pectin extracts. *Food Chemistry,* 100(4), 1356-1364.

[75] Zheng, Y., Lee, C., Yu, C., Cheng, Y. S., Zhang, R., Jenkins, B. M., VanderGheynst, J. S. (2013): Dilute acid pretreatment and fermentation of sugar beet pulp to ethanol. *Applied Energy,* 105, 1-7.

[76] Zheng, Y., Yu, C., Cheng, Y. S., Lee, C., Simmons, C. W., Dooley, T. M., Zhang, R., Jenkins, B. M., VanderGheynst, J. S. (2012): Integrating sugar beet pulp storage, hydrolysis and fermentation for fuel ethanol production. *Applied Energy*, 93, 168-175.

doi: 10.4028/www.scientific.net/KEM.548.39

Potential of Medium to More Fractured Natural Stone Deposits

Ivan Tomašić[a], Zrinka Vidović-Tisanić[b]

Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb,
Pierottijeva 6, 10000 Zagreb, Croatia

Exploitation and processing of small stone blocks and tombolons has a major impact on profitability. It is pointed out that the natural stone needs to be rationally used in deposits. Resources of rock mass tend to be greatly increased due to the utilization of different sizes of blocks in deposits or quarries.

This approach encourages the application of Lasky's relationship in the deposits of natural stone and quarries. Based on this relationship we can simply say that if the sizes of stone blocks and other stone products decrease arithmetically, then reserves increase geometrically or exponentially. On the presented examples, the utilization of the rock mass was analysed by the application of two different technologies of exploitation.

Besides, if the exploitation of stone blocks of different sizes achieves high utilization of the rock mass, then exploitation of aggregates and fillers increases much more. It is necessary to emphasize the existence of the border area based on favourable discontinuity density and spatial discontinuity position. Decision should be made whether to apply I or II variant of the exploitation (making vertical and horizontal cuttings or separation of the blocks along natural discontinuities, or using slanting cuts). The example is shown and analysed as a two-dimensional problem.

Key words:

Fractured Rock Mass, Lasky's Law, Natural Stone, Stone Blocks.

* The article was originally published in the journal „Key Engineering Materials" 584 (2013); 39-47 by the publisher Trans Tech Publications. The article has been approved for scholarly and non-commercial use. The reprint permission granted by the authors.
[a]ivan.tomasic@rgn.hr, [b]zrinka.vidovic-tisanic@rgn.hr

1. Introduction

Economic crisis in Croatia during the years 1964-1965 strongly influenced the exploitation and natural stone processing. The economic factors significantly decreased the number of active quarries and natural stone varieties on the domestic market. The size of the commercial blocks and the cost of labour generally caused the increase of cost because of exploitation with old and inappropriate equipment. It was the main reason for the closing of many quarries, especially those where small stone blocks dominated. After that period their number had been slowly increased for a long time.

The situation has significantly changed nowadays. Technology of exploitation and processing of small blocks and tombolons has been improved today. Stone blocks, mostly of smaller size, can be economically used and processed today. It mainly refers to the deposits of medium to more fractured rock mass. Small blocks and tombolons are pieces of stone smaller than 1.5 m^3. High decorative, physical and mechanical properties of small stone blocks and tombolons have also great positive influence on possible profitable exploitation and processing.

In this article it is presented how small blocks and tombolons could be economically exploited and processed in certain circumstances today. The assumption that permanent increase of discontinuity density has great influence on the utilization of rock mass in deposits is verified. The problem is analyzed in two dimensions. It is determined that the exploitation and processing of the small stone blocks and tombolons increase the available reserves due to rational utilization. It is also assumed that increase of discontinuity density in m^1/m^2 (in real conditions in m^2/m^3) causes the arithmetic decrease of stone blocks of specific sizes (in our case stone surfaces). If the exploitation of different stone block sizes achieves higher utilization of the rock mass, than exploitation of aggregates and fillers increases utilization of rock mass much more. It also enables cumulative increase of stone resources. This approach shows great similarity with the Lasky's law.

2. Probable Reserves of Natural Stone in Accordance to the Lasky's Relationship

If the small stone blocks and tombolons could be multiplied and additionally processed the added value should increase while the benefits and effectiveness of such

production would be higher. Production costs will be also slightly higher. This approach leads to enhancement of the usability of the rock mass in deposits or quarries of natural or dimension stone. The question is where are the possible limits of the total utilization of reserves? The limits of total utilization of reserves could be detected in the future regarding the potential of natural stone, stone blocks, tombolons, aggregates and fillers that we plan to produce. In particular, this means that Croatian natural stone of calcareous composition mostly limestone may be further rationally and sustainably exploited.

There are many legislative restrictions in Croatia, especially environmental, which complicate obtaining concessions for opening new exploitation fields with natural stone potential. High decorative, physical and mechanical properties of small stone blocks have also positive significant influence on market prices and efficiency of exploitation.

Interdisciplinary research of mineral resources with appraisal techniques was recognized by Samuel G. Lasky, a U.S. Geological Survey geologist, more than 60 years ago. His efforts to devise an estimate technique that incorporate many attributes of mineral resources resulted in cumulative tonnage-grade relationship with the average grade that has been referred to "the arithmetic- geometric (A/G) ratio" as an example of a mathematical model for estimating metal reserves [3].

This model was named after him the Lasky's law and states that if the average grade of metal content of ore in mine (expressed in percent per ton) decreases arithmetically, then the cumulative availability of ore increases exponentially. He formulated the equation of the relationship between average grade and cumulative tonnage as follows:

$$G = a + b \cdot \log T_o$$

where T_o is the cumulative tonnage of ore that is produced or obtained on the basis of estimates; G is the measured average grade of cumulative tonnage and a and b are constants determined for each deposit. Lasky has calculated the values of a and b according to the relative tonnage and medium contents of ore, or in some cases, the Cut-off grade (limit content) for the supply of various metals. Fig. 1 shows the relationship between the cumulative tonnage of ore and average grade (in percent per tons) according to Lasky's law and refers to copper ores available from porphyry copper deposits [3]. It demonstrates the construction of the Lasky's equation for 68 equally sized blocks of 25000.00 metric tons (t) of ore having average block grades of 0.2 percent; 0.4 percent; or 0.8 percent. The numbers are hypothetical and are chosen to simplify the example. In Croatia, the Lasky's law could be applied in estimating mineral reserves in the natural stone, oil and gas deposits.

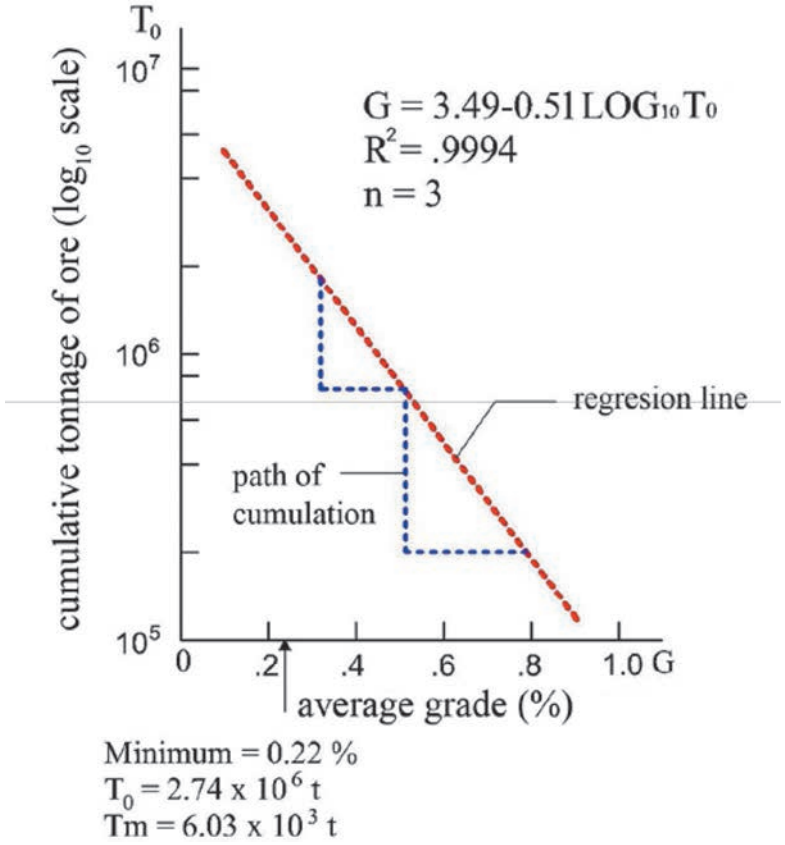


Fig. 1 – Relationship between cumulative tonnage and average grade, after Lasky [3].

Genesis and many postdiagenetic processes mostly affect the formation, distribution and accumulation of mineral raw material within a deposit. There are ore and mineral raw materials that have limited distribution and accumulation of reserves in the rock mass or within deposit itself. In that case, reserves will be, after exploitation, quickly and sharply reduced. Their appearance in profitable quantities will be completely limited and questionable. For such a deposit Lasky's law cannot be applied.

The practice has shown that the Lasky's law is applicable only in some deposits where the mineral component appears to have specific genesis and structure. The shape of ore body depends on a number of different geological factors. The primary form of the deposit could be completely changed due to displacements and later diagenetic processes. Mineral occurrences might be completely destroyed, changed and stretched over a wider area of the earth's crust by plication movements [3]. Such deposits offers hope that the mineral raw materials exist in large quantities. For such a deposit Lasky's law can be applied.

3. The Impact of Structural Fabric and Technology Exploitation on Utilization

To obtain a proper insight on the efficiency of the rock mass as a function of the structural fabric and applied technology during the exploitation, the whole problem is simplified and analysed as a two dimensional one. We give an example by calculating the coefficient of surface utilization of rock mass along the vertical stone walls or vertical sections in a quarry.

The model is based on discontinuities measured on the limited stone wall surface (6m x 6m) during the field research in the quarry of dolomite marble Sivec next to Prilep in Macedonia. Fig. 2a and Fig. 2b show the discontinuities intersected at an angle 80-82°, and with the dip of 44° and 56°. The figures show only two of many analysed examples of possible cross-section with different position and discontinuity density.

There are two variants of exploitation (I and II) which are shown in the analysed examples. The first one (I) refers to an example where the deposit exploitation allows the corresponding machines to make only vertical and horizontal cuts. The second (II) variant of exploitation involves performing the inclined cuts or separation of the rock mass with the appropriate machinery along the surface of natural discontinuity. Both variants of exploitation are shown in Fig. 2a and Fig. 2b. In Fig. 2a

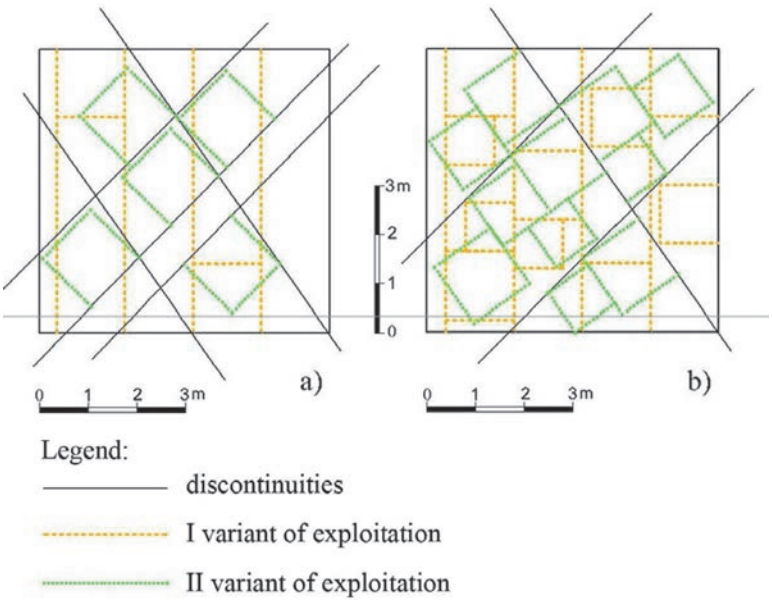


Fig. 2 – Hypothetical discontinuity density on the stone surfaces in the quarry during the two different variant of exploitation (I and II).

utilization is first estimated for surfaces of possible block dimensions of 1.4 × 1.4 m that could be exploited from the rock mass by vertical and horizontal cuts. This manner is used primary in healthy rock mass when the discontinuities are in a favourable spatial relationship and also at the increased distance. It yields high utilization and produce lower costs.

If the rock mass is intersected by less favourable positions of the discontinuities, the first (I) variant of exploitation gives low utilization of stone blocks (in this example block's surfaces) which should produce more waste while exploitation costs would grow. At the same time, it is shown in Fig. 2a that performing the inclined cuts gives a higher utilization of stone blocks (in this example block surfaces). The second (II) variant of exploitation (inclined cuts) gives 5 block's surfaces (Fig. 2a) in comparison with 2 block's surfaces (Fig. 2a) after using the first (I) variant of exploitation (horizontal and vertical cuts).

The exploitation (Fig. 2b) of three different blocks sizes (in our case block surfaces 1.4×1.4; 1.2×1.2, and 1.0×1.0 m) will significantly increase the utilization in both variants. In this example, the utilization shows a remarkable growth after the second (II) variant of exploitation (inclined cuts) in fractured rock mass. Great utilization of rock mass could be partly followed by increasing costs. Such access considerably increases the efficiency of utilization of more fractured natural stone deposits.

The problem presented and analysed in this way is similar to optimization problems in some other industries. For example maximum utilization of paper, tin or cloth is the same problem [1]. Also, the solution to this problem reminds of the results obtained by [4] for placing equal squares into larger squares of minimal surfaces.

The analysed examples shows that, instead of horizontal and vertical cuttings, utilization can be increased introducing inclined cuts in the quarry, in spite of unfavourable conditions of discontinuity density and discontinuity position. An inclined cut in rock mass is usually avoided. In this regard it is recommended, during the exploitation, to shift more fractured parts of rock mass on the base stage, where then it will be decomposed along the discontinuity.

The possibility of getting favourable large and small blocks from fractured parts of the rock mass is very important as is suggested in [2]. The authors suggest that much more fractured rock mass, without the use of explosives, should roll over or break off on a horizontal surface. After roll over, a great primary block should be totally or partially separate along natural discontinuities. This is followed by separation and formation of smaller blocks and tombolons along the discontinuities. Irregular stone blocks should be subsequently formed to market size and for man-

ufacturing. The advantages of these methods are justified by big productivity, reducing costs and raising the utilization and efficiency during the exploitation in the more fractured rock massifs.

It was emphasized [5] that the black powder use has a suppressing effect in more fractured rock masses. It is well known that explosives are not suitable for use during the exploitation of natural stone blocks. But, the black powder has slow combustion speed (400 m/s). In relation with other explosives it "gently" pushes fractured stone blocks on horizontal surface without any significant damaging and destruction. In this way, the minimum destruction of whole sections of the rock mass is achieved. After the roll over, the rock mass have to be separated along the natural discontinuities. Forming of the blocks of various sizes then begins. Increasing efficiency is similar to that described above in the second (II) variant of exploitation.

It was pointed out [6] that productivity and revenue depends on the size of the site, number of employees, investments and reserves, as much as on the skill of using available machines and tools. However, the importance of individual worker performance is emphasized. It also highlights the moderately intense exploitation of the quarry due to the utilization of rock mass of the deposit.

The utilization of stone blocks from rock mass in the natural stone quarry depends on the complexity of the rock fabric and on the application of appropriate technology exploitation [7]. Appropriate machine in the course of exploitation should be maximally adjusted to the jointed rock mass.

In order to assess how hypothetical technology of exploitation under different conditions of fractured rock mass can affect the utilization we used the analysis of numerous sections again. These cross sections were the basis for obtaining the curve shown in Fig. 3 and Fig 4, which clearly explain how the problem should be considered when choosing the method of exploitation. The ordinate shows the values of the discontinuity density (m^1/m^2), while the abscissa shows the values for utilization (%). The higher value of discontinuity density in the form of discontinuity length per unit area (m^1/m^2) means that the rock mass is more strongly fractured. High discontinuity density indicates the lower utilization of the great stone blocks from the rock mass. However, if we start with the exploitation of different sizes of stone blocks or in this case the block's surfaces, the utilization will be substantially increased. The working example has hypothetical and speculative character. This example is very instructive and can be well used in the educational sense.

The curves in Fig. 3 show the significant differences in the utilization of the rock mass between the first (I) and the second (II) variant of exploitation. It relates on stone blocks of constant surfaces 1.4 × 1.4 m. Differences could be higher if the

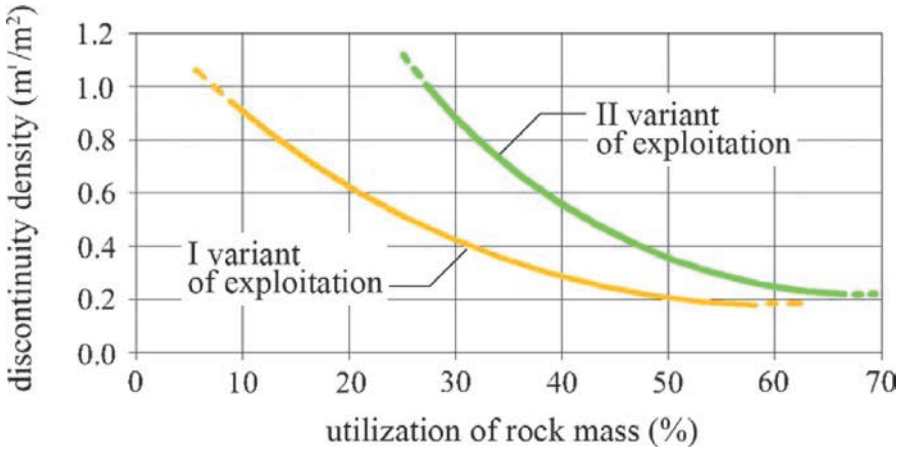


Fig. 3 – Curves showing the relationship of exploitation between discontinuity density and utilization of the block size surfaces 1.4×1.4 m [7].

rock mass is more fractured. That curves are based on the example shown before on Fig. 2a and the many other cross-sections used for analysis of utilization. By reducing the discontinuity density, the rock mass becomes less fractured and the differences in the utilization for both variants of exploitation are smaller.

The curves in Fig. 4 arise from the analysis of examples like the one shown in Fig. 2b and many others. Fig. 4 shows exploitation of stone blocks of various sizes (stone surfaces in our case) using I and II a variant of exploitation of the rock mass. Different sizes of their block's surfaces as is shown in this example on Fig. 2b are 1.4×1.4; 1.2×1.2 and 1.0×1.0 m. Both obtained curves are now significantly shifted to the right in comparison to the previous curves presented in Fig. 3. This refers to

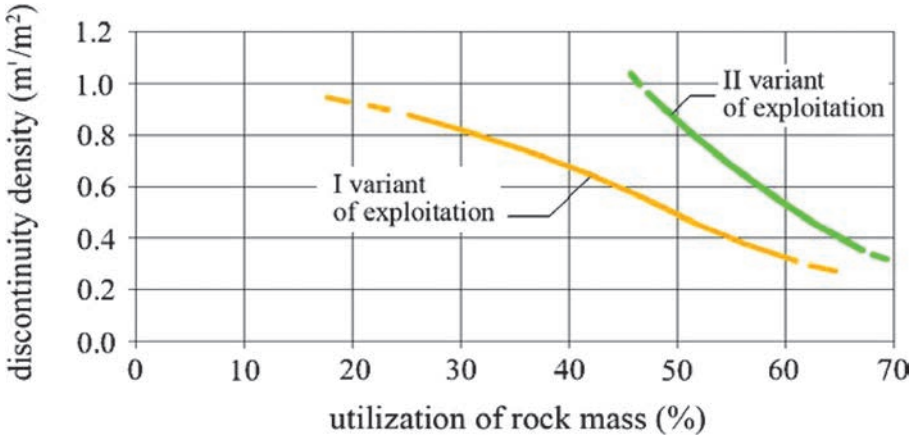


Fig. 4 – Curves of the relationship of exploitation between discontinuity density and utilization of block size surfaces 1.4×1.4; 1.2×1.2; 1.0×1.0 m [7].

both variants of exploitation (I and II) and suggests that the exploitation of different block sizes achieve greater efficiency. It is considered very favourable for exploitation.

This approach allows comparison with Lasky's law, but not entirely. Full approach to Lasky's law would have been achieved if the increase of discontinuity density is followed by reduction of block sizes. This means that, if we reduce the size of blocks (block's surfaces in our case) arithmetically, starting from sizes 1.4×1.4; 1.2×1.2; 1.0×1.0; 0.8×0.8; to 0.6×0.6 m or less because of increasing of discontinuity density, then the reserves of stone should grow geometrically or exponentially. Therefore, if the discontinuity density increases arithmetically the block size must be reduced arithmetically.

Therefore, the curves in Fig. 3 and Fig. 4, although clearly showing an increase in utilization, are not in accordance with the Lasky's law because three different surfaces of the analysed blocks were used for each of different discontinuity densities. The differences in the utilization of the rock mass between the first (I) and second (II) variant of exploitation are the greatest in more fractured rock mass (with the high discontinuity density).

There is no doubt, as it is shown here, that the three-dimensional representation of the utilization value is always lower than by the method applied here. It is of great importance to differentiate clearly the application of two essentially different variants of exploitation (I and II). Positive difference results from the exploitation of different dimensions of the blocks in relation to blocks of equal size (surfaces).

Therefore, there is no doubt, that the second version of exploitation shown in Fig. 3 and Fig. 4 in practice can achieve great utilization of rock mass. It is important to emphasize great tumbling down of rock mass (150-250 m^3) and its decomposition along natural discontinuities [2, 4]. Irregular blocks then could be formed into regular blocks or parallelepipeds. This method should be used for medium to more fractured rock masses.

This approach allows comparison with the Lasky's law. It means if the size of the stone blocks (surfaces) decreases arithmetically (reduction size of stone block corresponds to the reduction of average grade other mineral raw materials), then the reserve increases geometrically.

This suggests that in the future the development of stone building construction works will be mostly affected by proper application of technology exploitation and processing. In particular, this will affect the sustainable management of stocks of natural stone.

4. Concluding Remarks

Analysed examples showed that Lasky's law could be usefully applied in natural stone deposits. In this respect, the results of the analysis of the utilization of rock mass were suggesting the need for rational planning and managing of natural or dimension stone deposits.

In the medium and more fractured natural stone deposits, considering the rock mass density, application of appropriate technology exploitation should be applied.

Important decisions in the quarry should be made after the analysis of the measured discontinuities. Discontinuities and the discontinuity density are one of the most important structural fabric features that affect the size of the blocks that could be exploited. Its position and density could be measured and analysed in different ways [8].

Furthermore, analysed and presented results proved that the utilization of rock mass affects the proper selection of adequate technology of exploitation. Great discontinuity density of rock mass require exploitation of different stone block sizes. Exploitation of different stone block sizes may therefore be more profitable by the increment in utilization.

The view of the quarry and sawmill owned by the sculptors and entrepreneurs S. Vrandečić and A. Bogdanović on the island of Brač stresses the importance of the higher efficiency of the rock mass. Initial stage of development of a quarry is shown on Fig. 5. Stone blocks and tombolons smaller than 1,5 m^3 are visible in the foreground. Fig. 6 shows healthy and small stone blocks and tombolons ready for further cutting. Although smaller in size, all these pieces will be used for different purposes. In this way, the utilization of natural stone deposits will increase and sustainable development will be promoted.

The development of technologies not only maintains the viability of exploitation and processing, but also provides the possibility of increasing the efficiency of the rock mass of natural stone. Therefore, the growth of reserves could be possible. This means that the deposit will not be restricted only to the exploitation of large stone blocks. We will use those small stone parts from the deposit when the rock mass indicates a good quality. This is supported by a method that allows manufacturing of blocks and slabs, which are based on the epoxy treatment of stone defects.

In this regard, the biggest problem is: when to apply the first (I) variant of exploitation or the second (II) variant. It is obvious that in the healthy and low fractured rock massifs, horizontal and vertical cuttings (I variant of exploitation) allow ac-

Fig. 5 – View of the natural stone in the initial stage of quarry development.

Fig. 6 – Stone tombolons prepared for further manufacturing.

quisition of large blocks. The problem becomes much more complex if the discontinuity density increase.

There is also a border area where the position and discontinuity density open possibility to apply either I or II variant (making vertical and horizontal cuts, or separation of the rock mass along natural discontinuities or inclined sections). This possibility of choice significantly increases the efficiency, especially if taking into account the exploitation of smaller blocks. It can be assumed that in the border area the amount of overburden and costs of exploitation grows. This should be carefully analysed in medium and more fractured natural stone deposits.

The abscissa (Fig. 7) shows the fissured rock mass with different discontinuity density as shown in m^1/m^2. Discontinuity density increases to the left. On the same

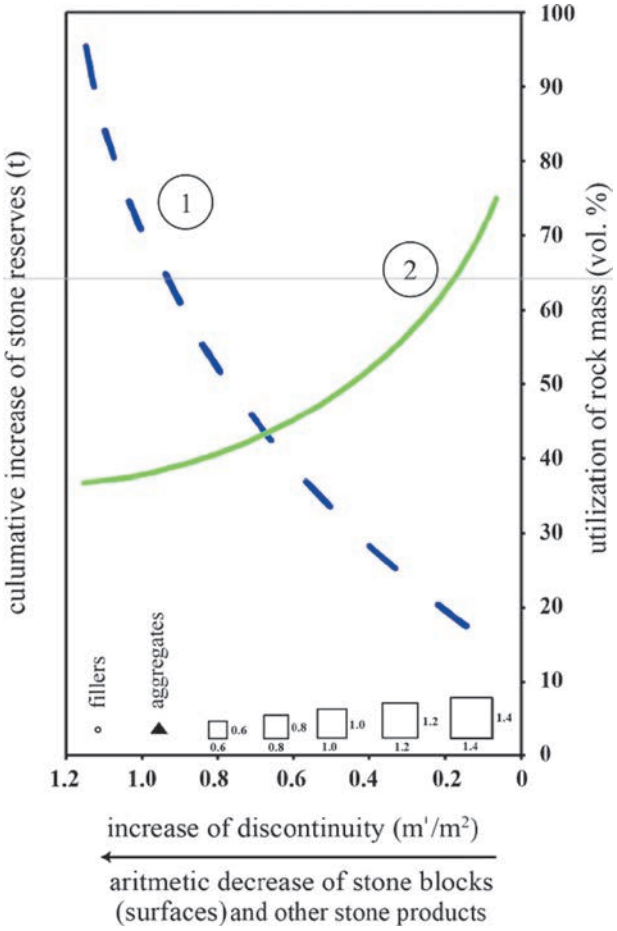


Fig. 7 – Hypothetical diagram (1) of possible cumulative increase of stone resources for different purposes and diagram of rock mass utilization (2).

diagram, the size of blocks reduces arithmetically in the same direction (in reality the volume of blocks). Hypothetical curve (1) shows the cumulative increase of reserves expressed in tons of natural stone in accordance with the Lasky's law (left ordinate). Curve (2) corresponds to the curve shown in Fig. 4 which is obtained by applying II variant of exploitation. In this case separation of the rock mass along natural discontinuities in the form of block's surfaces 1.4×1.4; 1.2×1.2 and 1.0×1.0 m is possible. The ordinate (right) shows the utilization of the blocks in vol. % (in our example the block surfaces).

The diagram indicates the importance of production of the different stone products with respect to size or weight. Scale of discontinuity density can be spread to the left, without loss of curvature of both curves.

All the results of the discontinuity analysis and utilization of stone blocks from the rock mass indicate that Lasky's law can be fully applied to the deposits of natural stone. This conclusion is based on hypothetical assumptions obtained on analysis and results shown in this article (Fig. 7). Because of specific relation, which is, that during the exploitation the size of the block reduces arithmetically in accordance with the increase of discontinuity density (in our example block's surfaces) we can achieve exact mathematical formula.

Deriving reliable exact method of prediction of exploitation it is important to notice that the Lasky's law, instead of the logarithmic form, can be written in exponential form:

$$T = c \cdot e^{kG}$$

or

$$T_o = c \cdot 10^{kG_o}$$

which is recognized as the solution of ordinary differential equations $dT/dG = kT$ that best establishes a connection between the arithmetic change of G's and exponential change of T 's. The constant k determines the prefix that decides if you are talking about exponential growth of T 's (positive k) or their exponential decline (negative k), and also determines the speed of change. The value of constants k and c depends on the characteristics of rock mass of each deposit and also determines the value of the ultimate points on the curve (1), (Fig. 7).

Curve (2) in Fig. 7 is derived from Fig. 4 and refers to the utilization of rock mass stone blocks in three different sizes of bocks (in our case the block surfaces 1.4×1.4; 1.2×1.2 and 1.0×1.0 m). It is in direct relation with the share of discontinuity den-

sity in rock mass according to the principle of exponential decline, and fall to the left (the higher discontinuity density the lower the utilization of stone blocks) for a specified stone blocks dimensions (surfaces).

The hypothetical curve (1) shows the cumulative increase of natural stone block resources in tons according to the Lasky's law. Accordingly, if discontinuity density increases, then block size reduces arithmetically while reserves (in m^3) grow geometrically, or exponentially. This means that at a certain discontinuity density more than one size of blocks should be exploited (in our case the surface). Besides the exploitation of stone blocks (expressed in tons) the rest of the stone should be used for the production of very small stone blocks, stone cubes, aggregates, fillers (in m^3 and tons) and stone for sculpturing. In this regard, the utilization of rock mass would be even more increased. Total reserves of stone would also be increased.

Finally it is necessary to say that only the usage of real block volume and other mentioned stone products can provide the actual curve (curve 1 in Fig.7). This approach will increase the value of deposits and increment in the revenue due to the added value, i.e. enabled with the great scale of different stone products. Some carbonate stone deposits (such as limestone in Croatia) are particularly suitable for this approach. It would enable sustainable exploitation and production of stone products. We believe that in the future these assumptions will be better validated and justified.

References

[1] A.R. Brown: Optimum Packing and Depletion. Mac-Donald–London and American Elsevier Inc., p. 107, New York. (1971).

[2] I. Cotman, and A. Damijanić: The New Excavation Method of Exploitation of Bedding Dimension Stone Deposits. The mining-geological-petroleum engineering bulletin, Vol. 3, p. 69-76, Zagreb. (1992).

[3] DeYoung, J.H., JR.: The Lasky Cumulative Tonnage-grade Relationship-A, Reexamination. Economic Geology, Vol. 76, p. 1067-1080. (1981).

[4] P.Erdös and Graham R.L.: On Packing Squares with Equal Squares. Journal of Combinatorial Theory 19, 119-123. (1975).

[5] K. Krebel: Method of exploitation of dimension stone by detonating cord and by black powder. Bulletin-Symposium: Dimension stone and aggregates, 167-176, Opatija. (1974).

[6] N.A. Milocco, Problemi di dimensioni di cave e miniere. Marmi-graniti e pietre, Vol 160, p 14-28, Milano. (1988).

[7] I. Tomašić: The utilization of building material deposits in dependence on the fabric and technology of exploitation. The mining-metallurgy bulletin, Vol. 29, p 332-342, Ljubljana. (1982).

[8] I. Tomašić, and A. Kršinić: Some Important Facts for Estimation of Natural Stone Deposits During the Exploration. World of Marmomacchine, International Multimedia yearbook, Promorama publisher, p 46-53, Milano. (2011).

doi: n/a

Logistics in University Tourism Curricula – EU vs. Croatia

Edna Mrnjavac[**]

University of Rijeka, Faculty of Tourism and Hospitality Management, Opatija

The aim of the paper is to indicate similarities and differences in university tourism curricula, regarding logistics. The research has been performed on EU university curricula, and the results have been compared with the situation in Croatia. The hypothesis has been defined as following: the presence of logistics in Croatian university curricula for tourism is less extended compared to the EU situation.

The purpose of the research is to establish the direction of curricula development in Croatia, in order to increase the level of compatibility with EU curricula, in the field of logistics. On the other hand, it is important to enable students to get enough cognition about logistics, for their implementation will lead to supply chain optimization and higher quality of tourism product.

Key words:

University tourism curricula, Logistics, EU, Croatia.

1. Introduction

Dynamic tourism development focused on numerous selective modes, which tend to establish themselves, on a fast-changing tourism market, by means of high – quality integrated destination products; seek highly educated and creative tourism profes-

[*] The article was originally published in the monograph „Logistics Flows Management in Tourism Destination“, 2012. by the publisher Faculty of Tourism and Hospitality Management, University of Rijeka. The article has been approved for scholarly and non-commercial use.

[**]ednam@fthm.hr

sionals. Tourism professionals' education plays a dominant role, therefore tourism development is in positive correlation with education development (Airey, 2002).

University institutions performing tourism curricula can be divided in two groups. The first one consists of those institutions that offer two curricula: tourism and hospitality. The second one is formed by institutions offering more curricula, with special reference to different elements in tourism. Logistics is in different ways present in both groups of institutions.

Logistics presence in tourism curricula has not been researched yet, despite the logistics role in tourism which has been recognised. When dealing with logistics of tertiary systems, notion and function of logistics systems in hospitality and tourism has been pointed.

Notion and function of logistics in tourism has been defined, logistics processes in tourism enterprises, as well as the concept of supply chain and management of logistics processes in tourism (Mrnjavac, 2010).

Being a main player in creating a tourism offer, elements of tourism destinations logistics have been researched, for efficient organisation of logistics flows in destination and in enterprises taking part in tourism offer, are a prerequisite for high-quality tourism product (Mrnjavac, 2002).

Logistics flows in tourism destination are researched in order to define main cause and effect reletionship and prerequisites to create an optimum logistics model (Baldigara and Ivanović, 2007; Mrnjavac and Ivanović, 2007). The object of research has been some selected logistics processes. This concerns to information processes (Galičić i Pilepić, 2007), but also to hotel supply processes (Mrnjavac, Pavia, i Kovačić 2009) as well as to organisation of aircraft catering (Ivanović i Vujić, 2007). For tourism services placement an optimalisation of logistics sales process is of utmost importance (Bašan i Berc-Radišić, 2007).

Among supply chain actors in tourism, a featured role is played by tour-operators, for their aim is to create a harmonised whole consisting of numerous services and products, based on logistics principles. Such a tourism product has to be market attractive (Muchina i Popovici, 2008).

Among types of selective tourism in which logistics plays an important role, event tourism has to be pointed, therefore processes and optimisation modes in event organisation have been investigated (Mrnjavac, 2005). Besides that a selected element of sport event is researched too, because its quality acts as travelling motive and reflects the attraction of tourism destination offer (Kovačić 2010).

Costs are also considered as object of logistics optimisation. It is upon them that the price of tourism product depends, but also its market attractiveness (Kordel, 2008).

Price is not the only criteria when deciding weather to buy a tourism product. Special-interest tourism opens numerous possibilities for hotel enterprises, while focus on selected market element seeks for approach adoption and process transformation in accordance to logistics principles. In such a way optimisation of involved resources leads to specialised hotel product, capable to guarantee desired quality for each special-interest guest (Kovačić 2011).

Hotel product quality aimed to special-interest market niche highly depends on the level of logistics concept incorporation in business and managerial functions, reflecting the correlation between logistics flows management and guests' satisfaction (Mrnjavac, Pavia, and Kovačić 2010). Hotel enterprise business model, where value for guest is based upon competences of employees, implies demand chain management, which includes internal marketing and supply chain management concept (Perišić and Kovačić 2010).

2. Defining Scientific Problem and Hypothesis about Logistics Presence in University Tourism Curricula

Logistics is primarily focused on flows optimisation which is present in production process, in order to produce a final product of quality characteristics which match the buyers needs. Logistics is focused on more and more flows. Special attention is dedicated to service logistics, with special reference to integral concept which tends to cover all elements and processes that participate in final product creation.

Tourism has to deal with harmonisation and management of different flows on all levels of system activities from micro to macro level, in order to create a complex product which will not only be acceptable to potential users but will also be much better compared to the competitive ones.

On a highly competitive tourism market a step towards lower costs and higher quality is considered a potential possibility to gain new service buyers. In this context the role of logistics is very important and still not researched enough and recognized.

With the growth of scientific interest about function and possibilities of tourism logistics, there are more and more logistics elements in tourism curricula. The presence of logistics in tourism curricula was not in the focus of research, although research about the role of logistics in tourism indicates acceptance and growth of interest for this area.

A scientific hypothesis has been defined: growing competitiveness on tourism market which encourages offer to create higher quality tourism product results in enlarged logistics presence not only in research but also in curricula.

It is presumed that the amount of logistics in tourism curricula in the Republic of Croatia lags behind logistics in EU curricula.

Expected scientific contribution consists in defining logistics elements that should by all means or at will be present in tourism curricula.

3. Analisys and Evaluation of Logistics Presence in EU Tourism Curricula

Material basis are European university tourism curricula. It should be emphasized that an attempt was made to take into consideration as much as European curricula as possible. Special attention was given to universities with tradition and high quality curricula. Such an approach was selected with the aim to analyze different educational concepts, conditioned by specific characteristics of surroundings in order to make a conclusion about universal ranges achieved despite those influences. These conclusions could indicate the direction of curricula development in Croatia.

The method of analysis was used to perceive European university curricula. Logistics elements have been selected and the frequency of its appearance defined. Comparison method was used to compare curricula according to title, number, duration and presence of courses. Method of synthesis related conclusions have been formulated, indicating certain rules, in order to define the compatibility of logistics contents with those in Croatian curricula.

By methods of induction and deduction from single cognitions about selected curricula a general cognition about logistics presence in curricula was created, in order to apply it on a model of curricula development in Croatia.

4. Analisys and Evaluation of Logistics Presence in EU Tourism Curricula – Research And Results

Research has covered three curricula on the theme of tourism logistics, but also tourism curricula which in their contents include selected knowledge in logistics.

In table 1 institutions performing curricula on the theme of tourism logistics are quoted.

Table 1 – Institutions performing curricula of tourism logistics

No.	Institution	Curricula
1.	Manchester Metropolitan University/UK	Logistics and Tourism
2.	College of Logistics /Ch.R	Logistics of Tourism
3.	RMIT University/Australia	Tourism Logistics

Source. Author's research 2012.

Manchester Metropolitan University from Great Britain offers curricula on the theme Logistics and Tourism on the undergraduate university level. College of Logistics from Prešov in Czech Republic is offering Tourism Logistics on undergraduate level as well, with the possibility to continue the study on graduate study in the field of logistics.

Notwithstanding it is not on the European ground the undergraduate curricula of Tourism Logistics is worth mentioning for in it some very important tourism fields, such as urban tourism, regional tourism, tourism market, information technologies, politics and planning, sustainability, event tourism, air transport etc. are being studied from the logistics point of view.

Unlike in table 1, table 2 contains institutions and curricula in which only selected logistics knowledge is present. The names of logistics courses are quoted.

Table 2 – Institutions and curricula containing only selected logistics courses, and course titles

No.	Institution	Curricula	Course title
1.	Brussels Business School	Tourism and Hospitality Management	Catering Management
2.	Modul University Vienna	International Tourism Management Tourism	Tourism production Chain Operations Management
3.	Cologne Business School	International Tourism Management	Sustainable Supply Chain Management
4.	University of Surrey	International Hotel Management	Travel Catering Management
5.	University of Lancashire	International Tourism Management	Managing Service Operations
6.	Northumbria University	Travel and Tourism Management	Travel Industry Management and Operations
7.	Manchester Metropolitan University	Events Management	Event Operations

Source: Author's research 2012.

Among 7 institutions the majority of them (4) are from Great Britain, one from Belgium, Austria and Germany. The title of even 3 curricula, from Austria, Germany and Great Britain is International Tourism Management. Only one program is devoted to Tourism and Hospitality, as well as to Management, Tourism, International Hotel management, Travel and Tourism Management and Event Management. One institution (Modul University Vienna), performs 2 curricula with selected logistics knowledge included.

Titles of courses are different. The majority of curricula (4) is focused on flows management under the name of: Operations Management, Managing Service Operations, Travel Industry management and Operations, Event Operations. In 2 curricula logistics is included in a course Catering Management and Travel Catering Management. In 2 curricula supply chain management is offered under the course title Tourism Production Chain and Sustainable Supply Chain Management.

When logistics is connected with some specific area it is usually done with: services, travel, event, supply, production and supply chains. In Croatia only Faculty for Management in Tourism and Hospitality within the curricula Tourism Management (Management of Tourism Destination) has recognised the importance of logistics in tourism and has included a course with a title Logistics of Tourism Destination, which deals with logistics in tourism and logistics management.

5. Logistics in European and Croatian Tourism Curricula – Discussion

5.1. Actual State in European Curricula

Research has indicated that there are reasons for logistics and tourism interdependence. Those reasons are being accepted in different ways, which is a consequence of developing needs of tourism in selected state, university law regulation, personnel availability and former university education development.

It could be deducted that logistics and tourism are mutually related in curricula, furthermore even 3 of them are named tourism logistics. Much more conceptually and not so many in contents curricula differ one from another. The Australian one is prominent for in each course tourism and logistics are combined – Urban Tourism Transport Logistics, Regional Tourism Transport Logistics, Information Technology and Tourism Logistics, Policy and Planning in Tourism Logistics etc., thus indicating the interdisciplinary basis on which curricula is based. High quality of lecturing is a difficult task, for teaching staff has to be familiar with both area –

logistics and tourism, which means that they themselves have to dispose of wide knowledge about logistics and tourism, in order to be able to define and explain the synergy between mentioned scientific disciplines.

British curricula of logistics and tourism do not show connectivity between two areas in each course, but only on the curriculum level. It includes knowledge about: actual economy, financial business, law, human potentials, economy and ecology, marketing and planning, creativity and entrepreneurship, tourism, tourism destination, tourism and sustainable development, management of tourism processes, quantitative methods in tourism and professional orientation in tourism. Such a concept can be explained by the fact that Manchester Metropolitan University offers a wide variety of tourism curricula, in which sometimes area without any connections to tourism are combined. More widely defined courses pave the way to more elasticity when choice of courses has to be done, regardless selected curricula, thus enabling to meet individual preferences of each student.

Czeck Curricula is conceptually similar to British one, because in encircles courses dealing or with logistics or with tourism. But it is more logistics oriented. Logistics is represented by courses like: Transport logistics, tourism logistics, logistics of passenger's travelling etc. As mentioned, students are allowed to continue studying on the graduate level in the field of logistics.

Among 3 curricula on the theme of logistics and tourism, another 8 tourism curricula on 7 university institutions, offer courses dealing with logistics. Basically those courses can be grouped in 3 groups:

- Managing operations (processes)
- Managing supply chain
- Managing catering

If these subjects are linked to a specific area, which is seldom done, it refers to travel and events.

It can be deducted that the accent is on processes management, business processes with the aim to increase efficiency, but also some parts of them – for example – production process.

Supply chain management, being one of the focuses of scientific interests in logistics research, plays an important role when tourism logistics is considered. Special attention is given to the business subjects that take part in supply chain organised by a tourism enterprise, which tends to integrate all parts of a business processes in one, to manage them all as a whole, so that all participants can fulfil planned and stipulated goals. For tourism enterprise the main task is higher quality tourism

product. Successful supply chain management results in logistics goals fulfilment for all business subjects involved in such integration.

Among logistics processes, hospitality processes differ from the others, because process of hospitality service production is very complex, made from numerous processes. Some of them are typical for service logistics, while the others for logistics of material products.

By harmonising procedures in these processes reduction of time, materials and work is obtained, tending to higher quality of product and lower production costs.

5.2. Actual State in Croatian Curricula

In Croatia logistics and tourism is being researched only on Faculty for Management in Tourism and Hospitality, within the curricula Tourism Management or Management for Tourism Destination. The course under the title Logistics of Tourism Destination includes basic logistics knowledge, tourism logistics according to the main tourism activities, concept of logistics of tourism destination and logistics management in tourism.

Although only one institution in Croatia has recognised the importance of logistics in tourism, and has provided for such knowledge in curricula, the situation with logistics in tourism in Croatia has to be positively validated, for it paves the way for other institutions that educate tourism manpower. It can be expected that others will follow the way that Faculty for Management in Tourism and Hospitality has marked.

Namely, other institutions that perform tourism curricula do not have such variety of courses, nor do they have enough lecturers and researchers capable of delivering lectures on desired professional level.

Concept of supply chain management is widely spread in tourism, although some modes of integration in tourism are not defined in such a way. It offers wide possibilities for cost cutting and quality augmentation. Therefore it can be expected that logistics will not remained limited only to one course, but will be spread to others.

Anyway the research results have verified the assumed scientific hypothesis that more and more spread competition on tourism market, which encourages suitable offer to create higher quality tourism product results in more extended application of logistics not only in research but also in curricula.

The second scientific hypothesis: that logistics in tourism curricula in Croatia is not so extended as it is in EU curricula, is only partially confirmed, because it is valid only for some tourism curricula. Namely, presence of logistics in Faculty for Management in Tourism and Hospitality curricula is on the same level as on EU institutions with only one logistics course. The majority of EU institutions (6 among 7 being included in research) perform only one logistics course.

Deviation in Croatian curricula can be deducted from the fact that there is no curricula Tourism logistics (or Logistics and Tourism), and that among 6 institutions offering tourism curricula only one has Tourism Logistics as a course.

Bearing in mind specific characteristics of Croatian tourism and education system, logistics contents that should be present in curricula are: basic elements of tourism logistics as service logistics, specific characteristics of tourism destination logistics, logistics management in tourism, specific characteristics of supply chain management regarding tourism enterprises-hospitality, agencies, transport, attractions bearers etc.

6. Conclusion

With the growth of scientific interest about function and possibilities of tourism logistics, there are more and more logistics elements in university tourism curricula. The presence of logistics in tourism curricula was not in the focus of research, although research about the role of logistics in tourism indicates acceptance and growth of interest for this area.

Research has covered three curricula on the theme of tourism logistics, but also tourism curricula which in their contents included selected knowledge in logistics.

Research has indicated that there are reasons for logistics and tourism interdependence. Those reasons are being accepted in different ways, which is a consequence of developing needs of tourism in selected state, university law regulation, personnel availability and former university education development.

Among 3 curricula on the theme of logistics and tourism, another 8 tourism curricula on 7 university institutions, offer courses dealing with logistics. Basically those courses can be grouped in 3 groups:

- Managing operations (processes)
- Managing supply chain
- Managing catering

If these subjects are linked to a specific area, which is seldom done, it refers to travel and events.

In Croatia logistics and tourism is being researched only on Faculty for Management in Tourism and Hospitality, within the curricula Tourism Management or Management for Tourism Destination. The course under the title Logistics of Tourism Destination includes basic logistics knowledge, tourism logistics according to the main tourism activities, concept of logistics of tourism destination and logistics management in tourism.

Research results have verified the assumed scientific hypothesis that more and more spread competition on tourism market, which encourages suitable offer to create higher quality tourism product results in more extended application of logistics not only in research but also in curricula.

The second scientific hypothesis: that logistics in tourism curricula in Croatia is not so extended as it is in EU curricula, is only partially confirmed, because it is valid only for some tourism curricula. Namely, presence of logistics in Faculty for Management in Tourism and Hospitality curricula is on the same level as on institutions with only one logistics course. The majority of institutions (6 among 7 being included in research) perform only one logistics course.

Bearing in mind specific characteristics of Croatian tourism and education system, logistics contents that should be present in curricula are: basic elements of tourism logistics as service logistics, specific characteristics of tourism destination logistics, logistics management in tourism, specific characteristics of supply chain management regarding tourism enterprises-hospitality, agencies, transport, attractions bearers etc.

References

[1] Airey, D. (2002). Growth and Change in Tourism Education, Rethinking of Education and Training for Tourism, Proceedings, Zagreb, pp. 13-22.

[2] Kovačić, N. (2010). Sports Event Logistics in Tourism, In Proceedings of the 20th International Congress Tourism and Hospitality Industry: New Trends in Tourism and Hospitality Management, May 6-8, Faculty of Tourism and Hospitality Management, Opatija, Croatia, pp. 976-988.

[3] Kovačić, N. (2011). Special-interest Management in Emerging Hospitality Brands, In Proceedings of the 30th International Conference on Organizational Science Development: Future Organization, March 23-25, University of Maribor, Faculty of Organizational Sciences, Portorož, Slovenia.

[4] Kordel, Z. (2008). Logistics in Tourism and Recreation, GeoJournal for Tourism and Geosites, Year 1, No. 2, Vol. 2.

[5] Mrnjavac, E., Pavia, N., Vujić, V. (2012). Interdisciplinarity in Higher Education Coursrs for Tourism: the Case of Croatia, Journal for Teaching in Travel and Tourism, 12, Routledge.

[6] Mrnjavac, E., Pavia, N., Kovačić, N. (2010). Special-interest Hospitality Logistics, Proceedings of 7th International Conference on Logistics and Sustainable Transport, Celje.

[7] Mrnjavac, E., Pavia, N., Kovačić, N. (2009). Implementation of Logistic Processes in Tourism, Proceedings of 6th International Conference on Logistics and Sustainable Transport, Celje.

[8] Muchina, S., Popovici, V. (2008). Logistics and Supply Chain Management in Tourism, Logistics of Merchandise, No. 24.

[9] Mrnjavac, E. (2010). Logistički menadžment u turizmu, Fakultet za menadžment u turizmu i ugostiteljstvu, Opatija.

[10] Mrnjavac, E. (2005). Logistics as Part of Event Tourism, 3rd Tourism Industry and Education Symposium, Jyvaskila Polytecnic School of Tourism and Services Management, Jyvaskyla.

[11] Mrnjavac, E. (2002). Logistics of Tourism Destination, Promet-Traffic-Trafico, FPZ, Vol. 14, No. 2, Portorož, Trieste, Zagreb.

[12] Perišić, M., Kovačić, N. (2010). Role of Internal Marketing and Logistic Process Management in the Hospitality Industry, In Proceedings of the 29th International Conference on Organizational Science Development: People and Organization, March 24-26, University of Maribor, Faculty of Organizational Sciences, Portorož, Slovenia.

[13] Zelenika, R. (2005). Logistički sustavi, Ekonomski fakultet, Rijeka.

[14] Zelenika, R., Pupavac, D. (2008). Menadžment logističkih sustava, Ekonomski fakultet, Rijeka.

doi: 10.1016/j.cmpb.2016.01.002

Multimodal Analysis of Startle Type Responses

Krešimir Ćosić[1], Siniša Popović[1,], Davor Kukolja[1], Branimir Dropuljić[1], Dragutin Ivanec[2], Mirjana Tonković[2]**

[1]University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, HR-10000 Zagreb, Croatia

[2]University of Zagreb, Faculty of Humanities and Social Sciences, Ivana Lučića 3, HR-10000 Zagreb, Croatia

Background and Objective: This article presents a multimodal analysis of startle type responses using a variety of physiological, facial, and speech features. These multimodal components of the startle type response reflect complex brain-body reactions to a sudden and intense stimulus. Additionally, the proposed multimodal evaluation of reflexive and emotional reactions associated with the startle eliciting stimuli and underlying neural networks and pathways could be applied in diagnostics of different psychiatric and neurological diseases. Different startle type stimuli can be compared in the strength of their elicitation of startle responses, i.e. their potential to activate stress-related neural pathways, underlying biomarkers and corresponding behavioral reactions.

Methods: An innovative method for measuring startle type responses using multimodal stimuli and multimodal feature analysis has been introduced. Individual's multimodal reflexive and emotional expressions during startle type elicitation have been assessed by corresponding physiological, speech and facial features on ten female students of psychology. Different startle eliciting stimuli like noise and airblast probes, as well as a variety of visual and auditory stimuli of different valence and arousal levels, based on International Affective Picture System (IAPS) images and/or sounds from International Affective Digitized Sounds (IADS) database, have been designed and tested. Combined together into

* The article was originally published by Elsevier in the journal „Computer Methods and Programs in Biomedicine".
Ćosić, S. Popović, D. Kukolja, B. Dropuljić, D. Ivanec, M. Tonković, Multimodal analysis of startle type responses, Comput. Methods Prog. Biomed. 129 (2016) 186-202.
The article has been approved for scholarly and non-commercial use by the publisher Elsevier.
**E-mail: sinisa.popovic@fer.hr; Tel.:: +385 1 6129837

more complex startle type stimuli, such composite stimuli can potentiate the evoked response of underlying neural networks, and corresponding neurotransmitters and neuromodulators as well; this is referred to as increased power of response elicitation. The intensity and magnitude of multimodal responses to selected startle type stimuli have been analyzed using effect sizes and medians of dominant multimodal features, i.e. skin conductance, eye blink, head movement, speech fundamental frequency and energy. The significance of the observed effects and comparisons between paradigms were evaluated using one-tailed t-tests and ANOVA methods, respectively. Skin conductance response habituation was analyzed using ANOVA and post-hoc multiple comparison tests with the Dunn-Šidák correction.

Results: The results revealed specific physiological, facial and vocal reflexive and emotional responses on selected five stimuli paradigms which included: (1) acoustic startle probes, (2) airblasts, (3) IAPS images, (4) IADS sounds, and (5) image-sound-airblast composite stimuli. Overall, composite and airblast paradigms resulted in the largest responses across all analyzed features, followed by sound and acoustic startle paradigms, while paradigm using images consistently elicited the smallest responses. In this context, power of response elicitation of the selected stimuli paradigms can be described according to the aggregated magnitude of the participants' multimodal responses. We also observed a habituation effect only in skin conductance response to acoustic startle, airblast and sound paradigms.

Conclusions: This study developed a system for paradigm design and stimuli generation, as well as real-time multimodal signal processing and feature calculation. Experimental paradigms for monitoring individual responses to stressful startle type stimuli were designed in order to compare the response elicitation power across various stimuli. The developed system, applied paradigms and obtained results might be useful in further research for evaluation of individuals' multimodal responses when they are faced with a variety of aversive emotional distractors and stressful situations.

Key words

Startle response; multimodal analysis; physiological, speech and facial features; airblast; International Affective Picture System; International Affective Digitized Sounds

1. Introduction

Startle stimuli represent various sudden and intense events that cause defensive responses, like a quick muscle contraction, activation of various facial and skeletal

muscles within a few milliseconds, leading to a whole body reflex that serves to protect the back of the neck and eyes. [1]. Three different components of the startle response should be distinguished [2]. First, muscle contractions protect vulnerable structures, like the neck or the eye ball. Beyond this immediate physical protection, a brief surge in activation of the autonomic nervous system (ANS) and endocrine response prepares the individual for a flight-or-fight response. This vegetative component of the startle response is reflected in a number of physiological signals, such as blood pressure [3], heart rate [3], skin conductance [4] and respiration [4]. Finally, ongoing cognitive processes are disrupted and attention is rapidly reoriented to the threatening stimulus [5;6].

Multimodality in startle type stimulation refers to a variety of visual stimuli, like images and video clips, and narratives with high arousal and application-specific associative contexts and semantics, which are superimposed on traditional startle stimuli. Such composite stimuli create more complex multimodal emotion elicitation with relatively high evoked response potentiation, referred to as elicitation power [7]. Such stimulation simultaneously targets several somatosensory cortex regions, including auditory, visual and tactile areas, and results in the broadest activation of different neural networks associated with potential traumatic events and experience. With such comprehensive semantically specific stimuli, a subject's potential deficits regarding his/her cognitive personal appraisal or hidden traumatic associative memories can be better assessed. Multimodality in the startle type response refers to reflexive and emotional reactions which can be captured by characteristic multimodal response signatures, like increased skin conductance, heart rate, contraction of the neck muscles, eye blink and laryngeal reflexive reactions, as well as emotional, cognitive and behavioral responses.

This paper presents an integrated multimodal analysis of responses to startle type stimuli via selected physiological, speech and facial features, unlike previous research works that typically focus on a single modality in combination with standard electromyogram (EMG) measurements of the eye blink startle response via the *orbicularis oculi* muscle. For example, in the physiological modality, Alpers and colleagues have conducted simultaneous analysis of *orbicularis oculi*, heart rate and electrodermal activity [8]. In the speech modality, the fundamental frequency (F_0) startle response analysis was proposed and studied by Sapir and colleagues [9], who compared F_0 to invasive EMG measurements of the laryngeal muscle response. In the facial modality, there have been a few studies of startle responses focused on automated eyelid detection and tracking in high-speed videos [10;11], as well as motion tracking of head and upper-body startle response with comparison to EMG measurements of the *orbicularis oculi* [12]. In comparison with previous papers, startle responses in this paper were analyzed simultaneously via dominantly expressed physiological, speech and facial features, i.e. skin conductance, speech fundamental frequency, speech energy, eye blinks and head movement. Additionally,

we have designed and evaluated a series of elicitation paradigms for startle type responses using visual, auditory, or tactile stimuli, as well as their combination, which has not been done previously.

2. The System for Multimodal Analysis of Startle Type Responses

The concept of the system for generation of various startle type stimuli and multimodal analysis of associated startle type responses by measuring physiological, vocal and facial reactivity is shown in Figure 1. The system is built around Lenovo S30 Workstation with Intel Xeon Processor E5-2620 v2 (15M Cache, 2.10 GHz), 32 GB of memory, and nVidia Quadro K200D graphics card, running Microsoft Windows 8.1 Pro operating system. Four system displays are related to: a display for supervisor system management, such as stimuli specification and specification of channels for multimodal data acquisition; two displays for stimuli presentation to the participants as well as experimenters; and a display for real-time visualization and tracking of selected multimodal response features like skin conductance, speech energy and fundamental frequency, eye blink, head translation etc. Systems for multimodal data acquisition include: Nexus-10 mark II device with 10 channels including skin conductance, electrocardiogram (ECG), respiration, skin temperature, blood volume pulse, EMG, as well as Biopac MP150 unit with ECG100C, RSP100C, EMG100C, GSR100C and SKT100C modules for physiology acquisition; voice acquisition system headset Sennheiser PC 360 and

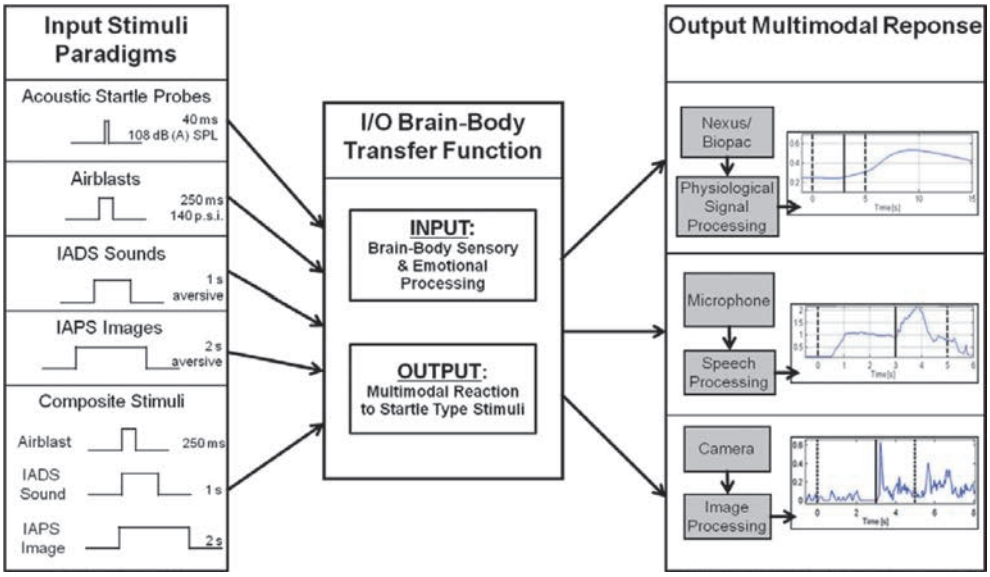


Fig. 1 – The concept of the system for input-output analyses of multimodal startle type responses.

throat microphone Telex TM750R; video acquisition system Logitech C920 web-cam.

The basic building blocks for stimuli generation are:

- The acoustic startle module which generates 108 dB (A) Sound Pressure Level (SPL) 40 ms white-noise bursts with 0 ms rise and fall times, delivered by a Samson audio amplifier and Sennheiser PC 360 professional audio headphones.
- The airblast delivery module [13;14] which includes Messer air tank, solenoid valve, camelback with nozzle attached to hose and solenoid controller. The 140 p.s.i. airblast is triggered by Measurement Computing DIO24 PCI Card.
- Elicitation of emotion based on images and/or sounds [15] which includes a variety of audio and visual databases like International Affective Picture System (IAPS) [16], International Affective Digitized Sounds (IADS) [17], NimStim Face Stimulus Set [18] and Nencki Affective Picture System [19].

Short highly arousing and negatively valenced visual stimuli and sounds may potentiate the emotional effects of acoustic startle elicitation paradigm. To enhance power of response elicitation in composite paradigms, highly arousing aversive IAPS video images and IADS sounds can be combined with 250 ms 140 p.s.i. airblast probes delivered to the participant's occiput. Additional aversive stimuli that can be delivered include highly irritating but not painful 500 ms electric shocks ranging from 0.5 to 5.0 mA, using electrodes attached to the fourth finger of the non-dominant hand. The goal of multimodal emotion elicitation is to provide a variety of stimuli, which are capable of eliciting broadband emotional startle type responses, in order to analyze the participant's vulnerabilities or robustness to complex aversive visual, auditory and tactile distractions.

3. Experimental Setting

Multimodal reactivity to startle type stimuli has been analyzed through five test paradigms which were designed and presented to ten female students, whose multimodal responses were captured via skin conductance, speech fundamental frequency and energy, as well as eye blinks and head movement. The paradigms contained established startle stimuli as well as their augmentation with the visual and auditory context and semantics which may elicit more complex multimodal physiological, facial and vocal responses, reflecting interactions of higher cognitive cortical structures, the brain stem and the emotional limbic system. For this purpose, IAPS images of unpleasant scenes, such as scenes of fear, disgust and mutilation,

were selected with congruent sounds. Specifically, each of the five test paradigms is described below:

1. **Acoustic Startle Probes**: Acoustic startle probes were simple white noise stimuli without any audio context and semantics. The acoustic startle reflex is elicited by auditory stimuli greater than 80 dB [20]; the current study used 108 dB.

2. **Airblast Stimuli**: Airblasts are aversive auditory and tactile stimuli, which also elicit startle response. Airblasts are usually directed to the larynx [21], but in this paper they were pointed to the occiput, due to interference with the participants' phonation and throat microphone (laryngophone) sensor.

3. **IADS Sounds:** Sudden and loud sounds based on IADS database were used as startle stimuli as well. These stimuli, unlike acoustic startle probes and airblasts, have specific semantics which additionally elicit emotional responses.

4. **IAPS Images:** Image context and semantics predominantly elicit emotional responses, rather than startle responses. Image-based emotional elicitation included a variety of visual databases, as described in the previous section.

5. **Composite Stimuli:** More real-world startle type stimuli included a combination of multimodal effects i.e. sounds, images, tactile sensing. Specifically, in this paper a combination of airblast with unpleasant IAPS/IADS images and sounds was used.

3.1 Participants

Ten female students recruited from the University of Zagreb, Faculty of Humanities and Social Sciences participated in the study. The selected participants were physically and mentally healthy, without any cardiovascular medical conditions, 18–19 years of age. All participants completed PTSD Checklist – Civilian Version (translated to Croatian), and their scores were below the cut-off level 30 for PTSD [22]. Two additional inclusion/exclusion criteria were used: the score on the emotional expressivity questionnaire [23] and the score on the anxiety sensitivity questionnaire [24;25]. Participants had to reach a high score on the emotional expressivity questionnaire (score $\geq$ 57) in order to select those with more intense emotional reactions. Assessment of students' anxiety sensitivity was added to eliminate participants for whom stimuli and physiological reactions (such as increased heart rate) could provoke panic attacks. Participants were required to have relatively low scores in the anxiety sensitivity questionnaire (score $\leq$ 40).

3.2 Procedure

After being screened for inclusion and exclusion criteria and signing the informed consent form in advance, on the day of the experiment the participants were familiarized with the laboratory setting. They were seated on a height-adjustable chair in front of a 75 cm-high table with 24" Lenovo ThinkVision monitor (1920 x 1200 resolution). The distance between the participant's eyes and the monitor was approximately 60–70 cm. Electrodes were attached and the participant was instructed to sit quietly, keep her eyes in the direction of the screen and neither move nor speak during the delivery of stimuli paradigms.

After the initial trials, participants underwent five elicitation paradigms (illustrated in Figure 2): (1) acoustic startle white-noise probes lasting 40 ms, with intensity of 108 dB (A) SPL; (2) airblasts to the occiput lasting 250 ms, using air pressure of 140 p.s.i.; (3) loud aversive IADS sounds lasting 1 s; (4) aversive IAPS images lasting 2 s; and (5) simultaneous combinations of aversive airblasts, IADS sounds and IAPS images with the same properties as in the prior respective paradigms. Each paradigm lasted 120 seconds and included three pseudo-randomly ordered time intervals with a stimulus presentation and three when stimuli were omitted. In order to capture the impact of stimuli on speech responses, the participants were given on-screen visual reminders to start sustained phonation of vowel 'a', immediately after onset of a white screen with a fixation cross '+'. All participants underwent the same paradigms in the same order, 1 through 5, with approximately three-minute breaks between successive paradigms. To evaluate the participants' perception of each elicitation paradigm, subjective ratings were collected regarding unpleasantness, stimuli intensity, overall self-assessed reactivity, and aversiveness ranking of the applied paradigms.

The experimenter verbally informed the participant regarding the beginning of each paradigm, and when the paradigm was finished. Each paradigm started with the black screen, which was followed by the first appearance of the white screen with fixation cross '+'. The white screen indicated that the participants should start sustained phonation of vowel 'a' until the black screen would reappear again 5 seconds later. During each paradigm lasting 120 seconds, there were 6 occurrences of the white screen with the fixation cross, which included in pseudo-random order: 3 No Stimulus (NS) conditions, in which the onset of white screen was only followed by reappearance of black screen 5 seconds later; and 3 Stimulus (S) conditions which included delivery of a specific startle type stimulus at t = 3 s after the onset of the white screen. The recorded physiological, speech and facial features were statistically analyzed in MATLAB for response differences between Stimulus and No Stimulus conditions within each paradigm, and response differences between the various paradigms.

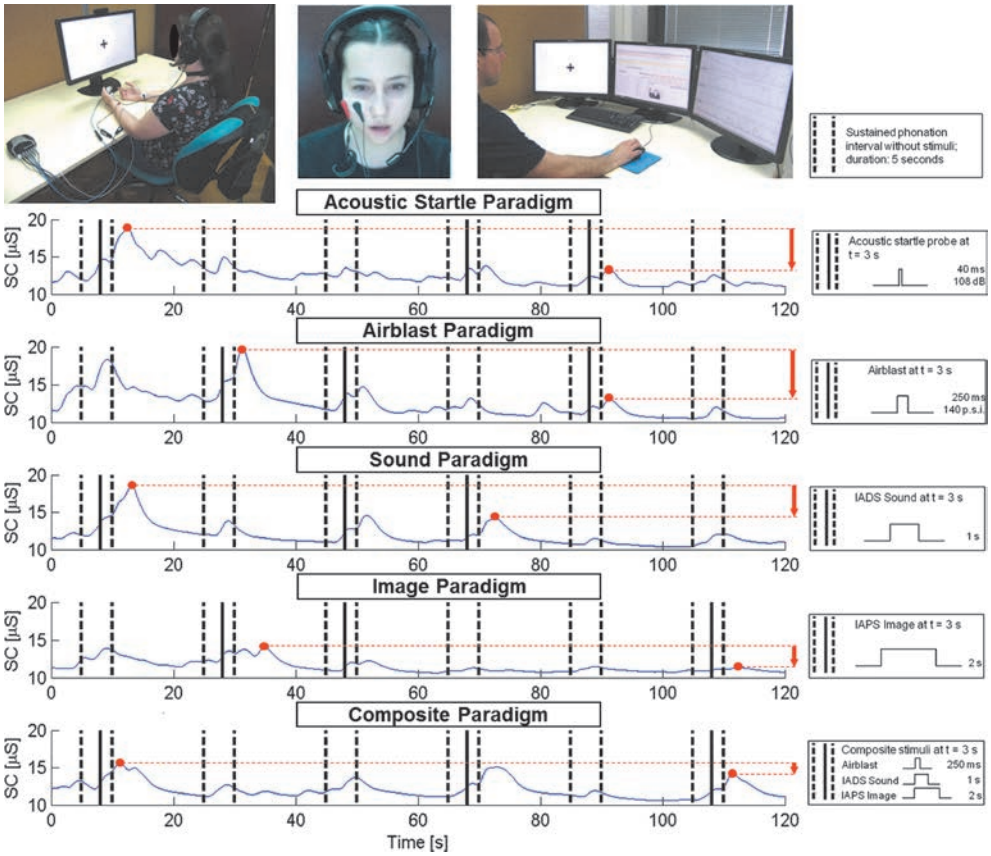


Fig. 2 – Illustration of skin conductance responses on five startle type stimulation paradigms. Two neighboring dashed lines reveal the phonation interval which was displayed on the monitor by fixation cross '+', while vertical solid lines reveal onsets of corresponding startle type stimuli. The blue curve illustrates the skin conductance level of a representative participant recorded during all five paradigms. The red dots mark the response peak to the first and last stimulus presentation within each paradigm, and the red arrow shows the degree of habituation.

The timing diagram of all stimuli paradigms, and skin conductance responses of a single participant on acoustic startle probes, airblasts, IADS sounds, IAPS images, and composite stimuli is shown in Figure 2. Figure 2 also illustrates the skin conductance habituation as a form of non-associative learning which decreases skin conductance response magnitude after repeated presentation of the same stimulus. It is also evident that habituation was stimulus specific [26] and different among all five paradigms. The habituation analysis was important because it can be disrupted in patients suffering from different mental health and neurodegenerative diseases [27]. Also, conditioned responses during the No Stimulus period can be observed in the increased skin conductance to the fixation cross which was associated with increased probability of the occurrence of an aversive stimulus.

4. Analysis of Multimodal Components of Startle Type Responses

Participants' behavioral reactions can be monitored in real time by selected physiological, speech and facial variables, like: skin conductance, fundamental frequency, energy, eye blink and movement of the participant's head. During the experimental measurements, all relevant multimodal input/output variables, sampling periods, time stamps, types of stimuli, time intervals of Stimulus and No Stimulus periods were continuously recorded for comprehensive off-line analysis. Therefore, any segmentation of recorded data could be easily examined in after-action analysis and review. More detailed analyses of selected physiological, speech and facial features are presented in the following sections.

4.1 Analysis of Skin Conductance Responses

Skin conductance level is one of the most robust non-invasive physiological measures of ANS [28]. It is used as a measure of arousal in many anticipatory anxiety studies and therefore was used as an important measure of startle type responses in our study. Laboratory studies have also found that skin conductance response varies linearly with arousal ratings regardless of the valence ratings [29].

Skin conductance and other physiological signals, like ECG, blood volume pulse, respiration, and skin temperature, were measured with Nexus-10 mark II physiology acquisition equipment wirelessly connected to the system by Bluetooth with a sampling frequency of 256 Hz for all channels. The physiology acquisition software was based on Nexus API and software for signal processing, feature computation and recording. Numerous physiological features described in [30] were calculated from physiological raw signals, and off-line analysis was conducted using MATLAB.

In the current study, startle type response was predominantly expressed using skin conductance level. Before feature computation, skin conductance signals were normalized on the basis of the minimum and maximum of the signal in all five stimuli paradigms, as recommended in [31]. To find minimum and maximum for each participant, we linked together skin conductance signals over the acoustic startle, airblast, sound, image and composite paradigms (Equation 1). From this data set, we computed normalized skin conductance signals (normSC) for each paradigm (Equation 2).

$$SC = \left[SC_{acoustic\_startle} \quad SC_{airblast} \quad SC_{sound} \quad SC_{image} \quad SC_{composite} \right] \tag{1}$$

$$normSC_i = \frac{SC_i - \min(SC)}{\max(SC) - \min(SC)} \qquad (2)$$

$$i = startle, airblast, sound, image, composite$$

The magnitude of the skin conductance response for the time interval after sustained phonation (5 s, 15 s) in paradigm *i* was calculated by the following equation:

$$magnitude(normSC_i) = \max(normSC_i(5\ s, 15\ s)) - normSC_i(0\ s)\,, \qquad (3)$$

where 0 s represents the onset of the white screen with fixation cross '+'.

Normalized skin conductance signals for each paradigm are shown in Figure 3, averaged across all participants in the Stimulus condition (10 participants x 3 stimulus

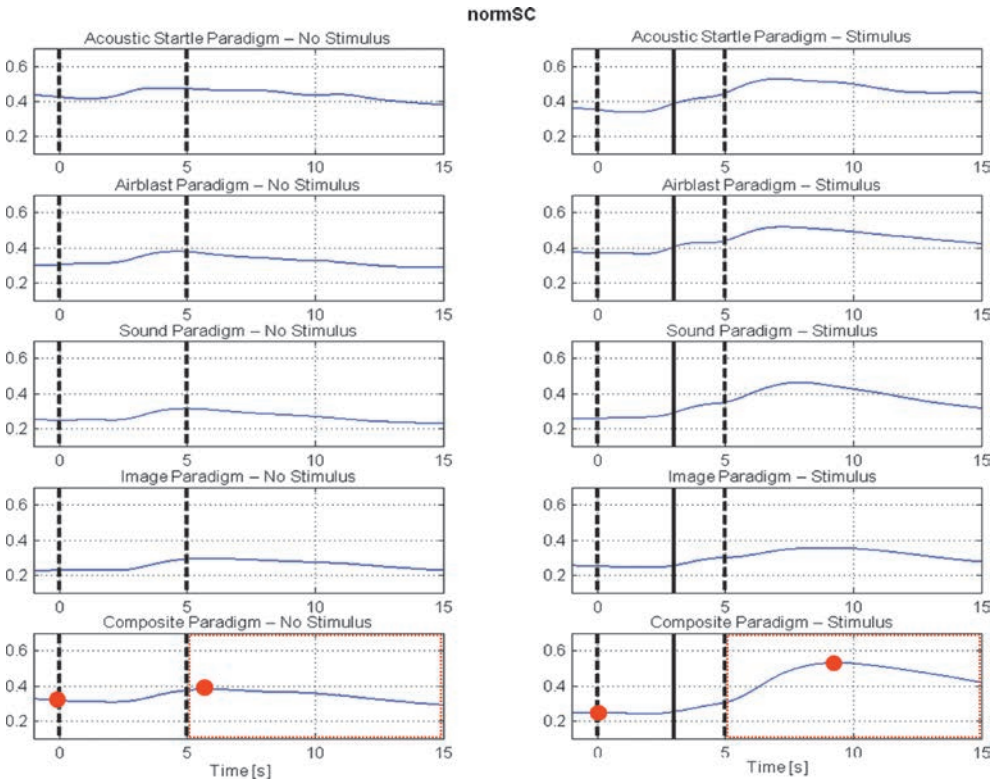


Fig. 3 – Average normalized skin conductance signals before, during and after vocalizations for trials without stimuli presented (No Stimulus condition, left) and for trials when stimuli occurred (Stimulus condition, right), across all participants.

trials) versus No Stimulus condition (10 participants x 3 trials without stimuli). As shown in Figure 3 the difference between skin conductance responses in the No Stimulus and Stimulus conditions is highest for the composite paradigm (red circles). Relatively high differences were also achieved with acoustic startle, airblast and sound paradigms, while minor difference was achieved with image paradigm.

Figure 4 illustrates the differences between the magnitude(normSC) calculated when startle type stimuli occured and the magnitude(normSC) when no startle type stimuli occur. In each paradigm, the magnitude(normSC) was computed from the participant's average normSC signal across 3 Stimulus conditions and, likewise, across 3 No Stimulus conditions. The greatest median differences occurred in the composite paradigm, followed by sound, acoustic startle and airblast paradigms, and, finally, image paradigm. Therefore, the median differences indicate that the composite paradigm had the highest response elicitation power for magnitude(normSC). It is also evident that, across all paradigms, participant 6 most consistently exhibited the lack of skin conductance responses.

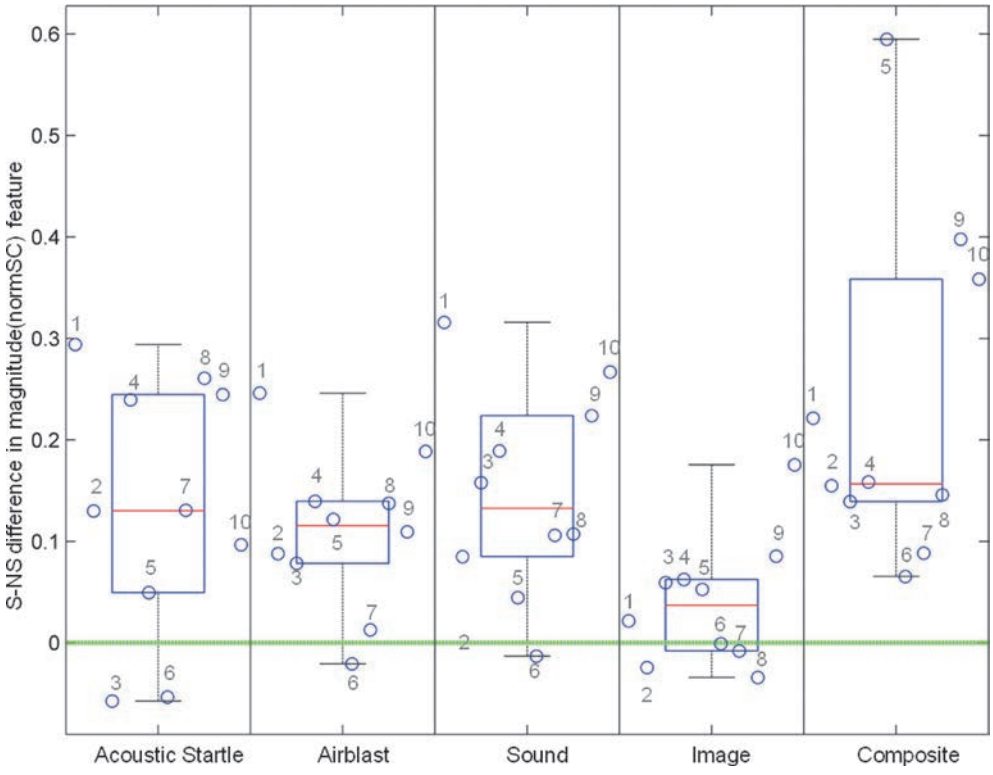


Fig. 4 – Distribution of the differences between each participant's average magnitude(normSC) feature in Stimulus (S) and No Stimulus (NS) conditions, for each paradigm. Numbers 1-10 indicate individual participants.

Additionally, to take into account the variability of responses between participants, an alternative measure of response elicitation power was computed for each paradigm from the individual participants' data plotted in Figure 4. This alternative measure was the effect size, computed as Cohen's d_z [32]:

$$\text{Cohen's } d_z = \frac{M_{diff}}{\sqrt{\frac{\sum_{i=1}^{10}(X_{diff,i} - M_{diff})^2}{N-1}}} \tag{4}$$

where the numerator represents the mean (M_{diff}) of the individual participants' average differences ($X_{diff,i}$) between Stimulus and No Stimulus conditions, which is being compared to zero, while the denominator represents the standard deviation of these differences. The results are shown in Table 1. The ranking of paradigms according to Cohen's d_z measure of the power of response elicitation is: (1) sound – highest; (2) airblast; (3) composite; (4) acoustic startle; (5) image.

Table 1 – Effect size of startle stimuli on the magnitude(normSC) (Cohen's d_z, N = 10).

Acoustic startle	Airblast	Sound	Image	Composite	Average
1.050	1.415	**1.453**	0.621	1.392	1.186

After applying one-tailed one-sample t-tests to the data from each paradigm in Figure 4, the greater-than-zero effects of stimuli were statistically significant ($p < 0.05$) for all paradigms. Repeated measures ANOVA rejected the null hypothesis of no difference in responding to various stimuli at 0.05 level of significance ($p = 0.0098$).

Skin conductance responses to startle type stimuli were also tested for habituation, by comparison of the magnitude(normSC) between the first, second and final third occurrence of startle type stimulus in each paradigm. Generally speaking, participants habituated to stimulation within the paradigms as shown in Figure 5. It can be seen that there was a continuous negative trend in responses within all paradigms, except for composite stimuli. Repeated measures ANOVA found statistically significant ($p < 0.05$) differences among responses to the first, second and final third occurrence of startle type stimulus in the acoustic startle, airblast and sound paradigm, but not in the image and composite paradigm. Post-hoc pairwise comparisons using Dunn-Šidák correction showed that there were significant differences between responses on the first and second, and the first and third stimulus in the acoustic startle and airblast paradigm, and that there were significant differences between responses on the first and third stimulus in the sound paradigm. There

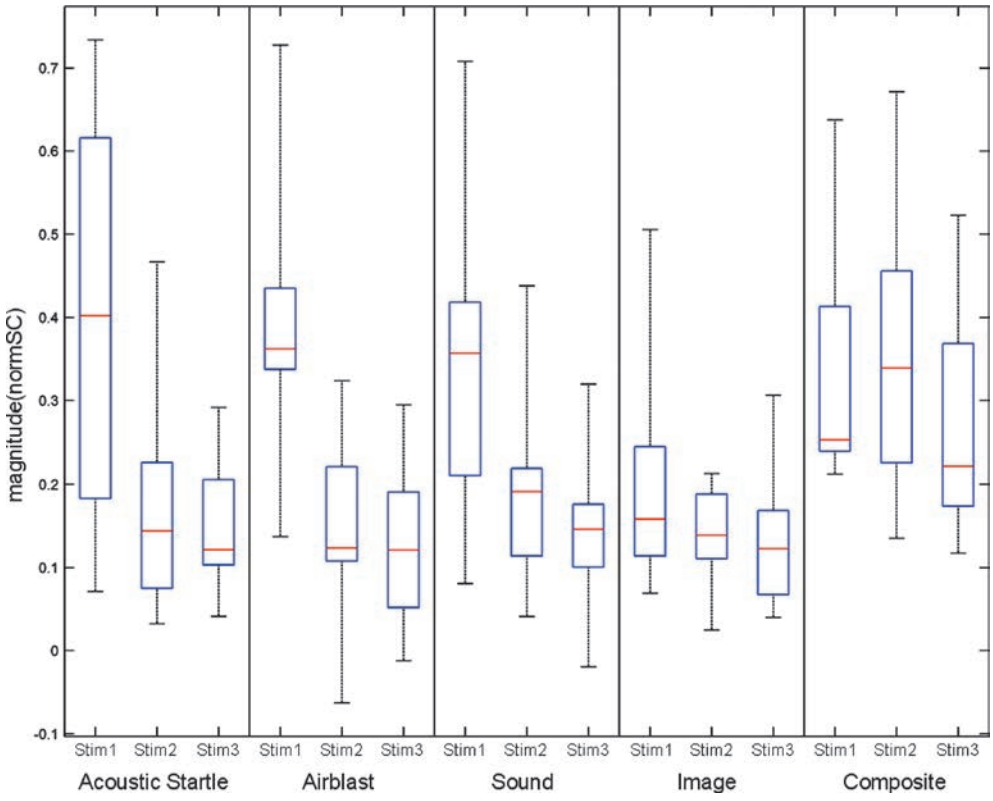


Fig. 5 – Distribution of the magnitude(normSC) after the first, second and third stimulus, for each paradigm. Copyright © 2016 Elsevier Ireland Ltd. http://www.sciencedirect.com/science/journal/01692607

were no significant differences between responses on the second and third stimulus within any paradigm.

4.2 Analysis of Speech Energy and Fundamental Frequency

The startle response in laryngeal muscles has not been explored as much as the eye blink response, but a few studies have shown that EMG response can be recorded from several laryngeal muscles when presenting startle stimuli. For example, EMG response was measured on cricothyroid muscle of the larynx during acoustic startle probe [9]. Contractions of the cricothyroid muscle elicited by an acoustic startle probe cause the reflexive vocal folds elongation, which consequently results in rapid short-term increase in fundamental frequency (F_0) response magnitude. Other studies somaesthetically elicited EMG responses on the *laryngeal vocalis* and inter-arytenoid muscles that are similar to those of the *orbicularis oculi* auditory startle

response [33], while similar results for laryngeal muscle innervations in the case of an airblast [34], and electrical stimuli [35] are also presented.

The main advantage of acoustic analysis of startle responses vs. EMG analysis of laryngeal muscles is in the fact that it is more practical; however it requires sustained phonation from the participant during the startle elicitation.

The F_0 startle response analysis was in its basic form proposed by Sapir and colleagues [9]. It was compared to laryngeal muscle response (the cricothyroid muscle), which was measured with invasive EMG electrodes. Stimuli intensity used in this research was in the range of 25 to 85 dB (A) SPL, but according to the 2005 eye blink guidelines [36], the recommended intensity of acoustic startle stimulus is 100 dB (A) SPL or higher. Also, in [37] was shown that eye blink occurs only in about 50% cases when intensity of a startle stimulus is set to 85 dB (A) SPL. For EMG eye blink reaction, generally an increase in the intensity of stimulus causes an increase of the response magnitude and the probability of a reaction, as well as a decrease of response latency [38]. In our previous work [39;40], acoustic startle stimuli were distributed in the range of 55 to 105 dB (A) SPL and similar conclusions on F_0 response were found. In this experiment, stimulus intensity was set up to 108 dB(A) SPL and F_0 analysis of the startle type response was extended with an energy (intensity) analysis.

Real-time computation of speech features like F_0 and energy was done by openSMILE v2.0 toolkit [41]. Sampling frequency of the raw speech signal was 8 kHz, monaural, with 16 bits per sample. Fundamental frequency and energy of speech were computed using a frame-based analysis (hamming window) where frame size was set to 25 ms and frame step to 10 ms. During initial testing, it was observed that sounds from headphones and airblast were recorded by the headset microphone, which would preclude analysis of the participant's speech features for airblast, sound and composite paradigms. Therefore, Telex TM750R throat microphone (laryngophone) was used to capture only the participant's raw speech signal.

Fundamental frequency was measured using ACF (autocorrelation function)/Cepstrum based method obtained directly from openSMILE v2.0 toolkit. Sustained phonation of the vowel 'a' by each participant provided fundamental frequency approximately as a steady state F_0 contour. Therefore, any change in the contour that occured as a response to the stimuli presentation could be easily detected. We assumed that the time interval (1.5 s, 3.0 s) after the onset of the white screen with fixation cross '+' represented a clear phonation, since in this period the influence of the initial breath and the turbulence at the beginning of the phonation were mainly eliminated. The response interval was taken immediately after the stimulus onset, i.e. (3.0 s, 3.5 s). The magnitude of F_0 for each phonation was calculated as a ratio of maximum F_0 values in time intervals (3.0 s, 3.5 s) vs. (1.5 s, 3.0 s):

$$magnitude(F_0) = \max\left(F_0^{(3.0\,s,3.5\,s)}\right) \Big/ \max\left(F_0^{(1.5\,s,3.0\,s)}\right) \quad (5)$$

Energy (intensity) of speech signal was measured using a root-mean-square method obtained from openSMILE v2.0 toolkit. In this case, a slowly decreasing contour was expected during the phonation, and its changes due to stimuli presentations were relatively easy to detect. The baseline interval of speech energy contour was (1.5 s, 3.0 s), while the response interval was (3.35 s, 4.5 s), i.e. 3.35 s after the onset of the white screen and 0.35 s after the stimulus onset in the case of stimulus event. The first 350 ms after the stimulus onset were avoided due to transmission of the airblast sound to throat microphone in a few cases. In these paradigms, airblast was presented within the first 250 ms after the stimulus onset, while an additional 100 ms (350 ms in total) were added to exclude all possible delay effects. The magnitude of energy (E) for each phonation was calculated as a ratio of maximum E values in (3.35 s, 4.5 s) and (1.5 s, 3.0 s) time intervals:

$$magnitude(E) = \max\left(E^{(3.35\,s,4.5\,s)}\right) \Big/ \max\left(E^{(1.5\,s,3.0\,s)}\right) \quad (6)$$

Time segments of F_0 and energy for each paradigm are shown in Figure 6, averaged across all participants in the Stimulus condition (10 participants x 3 stimuli) versus No Stimulus condition (10 participants x 3 trials without stimuli). Responses are evident in the interval (0 s, 0.5 s) for F_0 and in the interval (0 s, 1.5 s) for speech energy E after the onset of startle type stimuli in all paradigms except the image paradigm (e.g., see red circles for illustration). It must be noted that all the utterances were taken into account despite the fact that some utterances contained halving and doubling effects of F_0, short discontinuities in F_0 contour after the stimulus onset (in the case of unvoiced speech or silence, followed by phonation again), as well as short drops to zero in energy contour after the stimulus onset (a short silence interval, followed by phonation).

Figure 7 shows speech F_0 and energy magnitude differences between the Stimulus and No Stimulus conditions. In each paradigm, both features were computed from the participant's average F_0/energy contours across 3 Stimulus conditions and, likewise, across 3 No Stimulus conditions. The largest median difference of the magnitude(F_0) occurred in the airblast paradigm, followed by the composite, sound, acoustic startle and image paradigms. The median differences indicated that the airblast paradigm had the highest response elicitation power for magnitude(F_0). It is evident that, across all paradigms, participant 8 most consistently exhibited F_0 responses to aversive stimuli, while participants 4 and 9 most consistently exhibited a lack of F_0 responses. On the other hand, the largest median difference of the magnitude(E) occurred in the composite paradigm, followed by sound, airblast, acoustic startle and image paradigms. In this case, the composite paradigm had the highest response elicitation power. Across all paradigms, participant 8 again most consistently exhibited responses to aversive stimuli, while no particular participant exhibited a lack of energy responses.

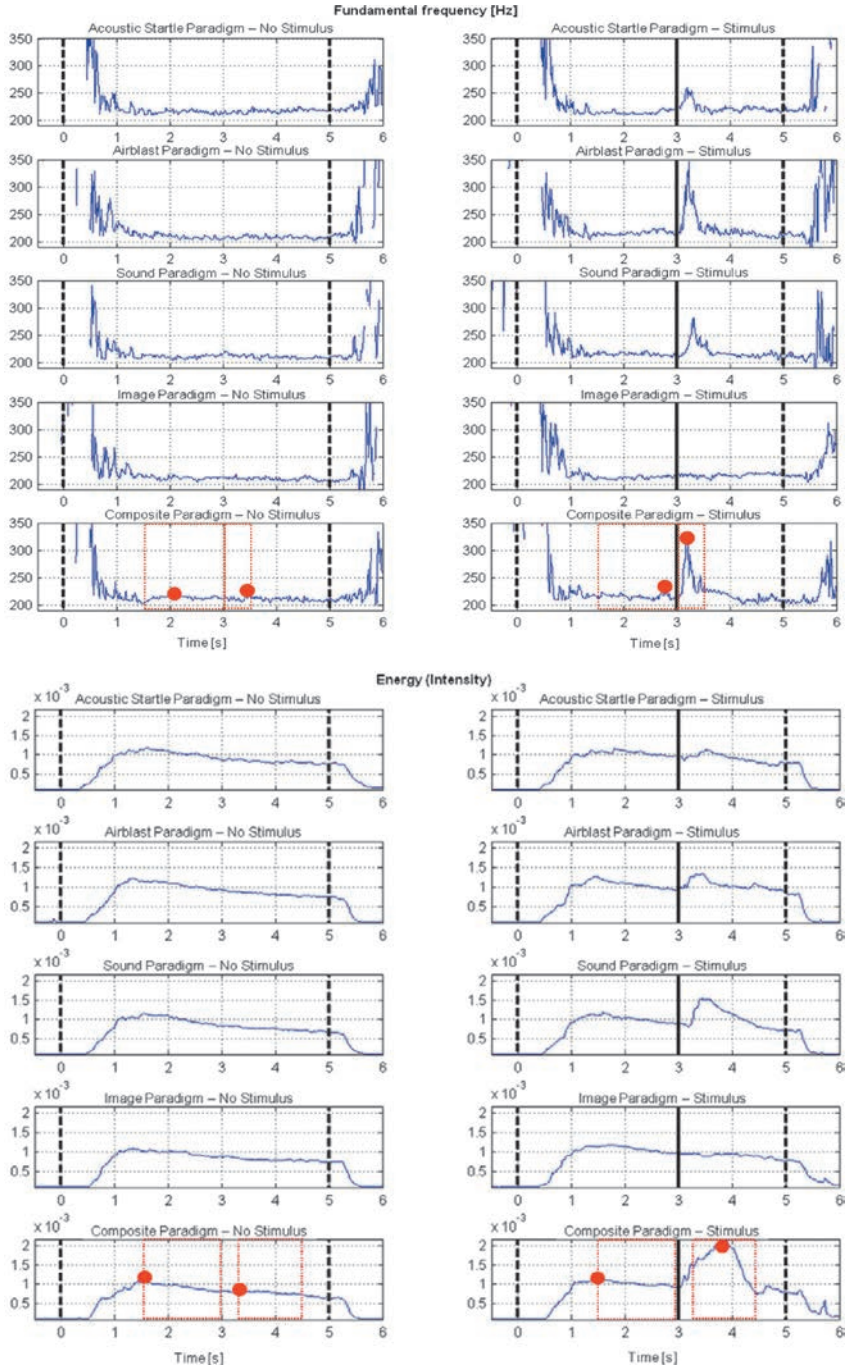


Fig. 6 – Average speech fundamental frequency (top panel) and energy (bottom panel) time series across all participants shortly before, during and shortly after phonations, for trials without stimuli presented (No Stimulus columns) and for trials when stimuli occurred (Stimulus columns).

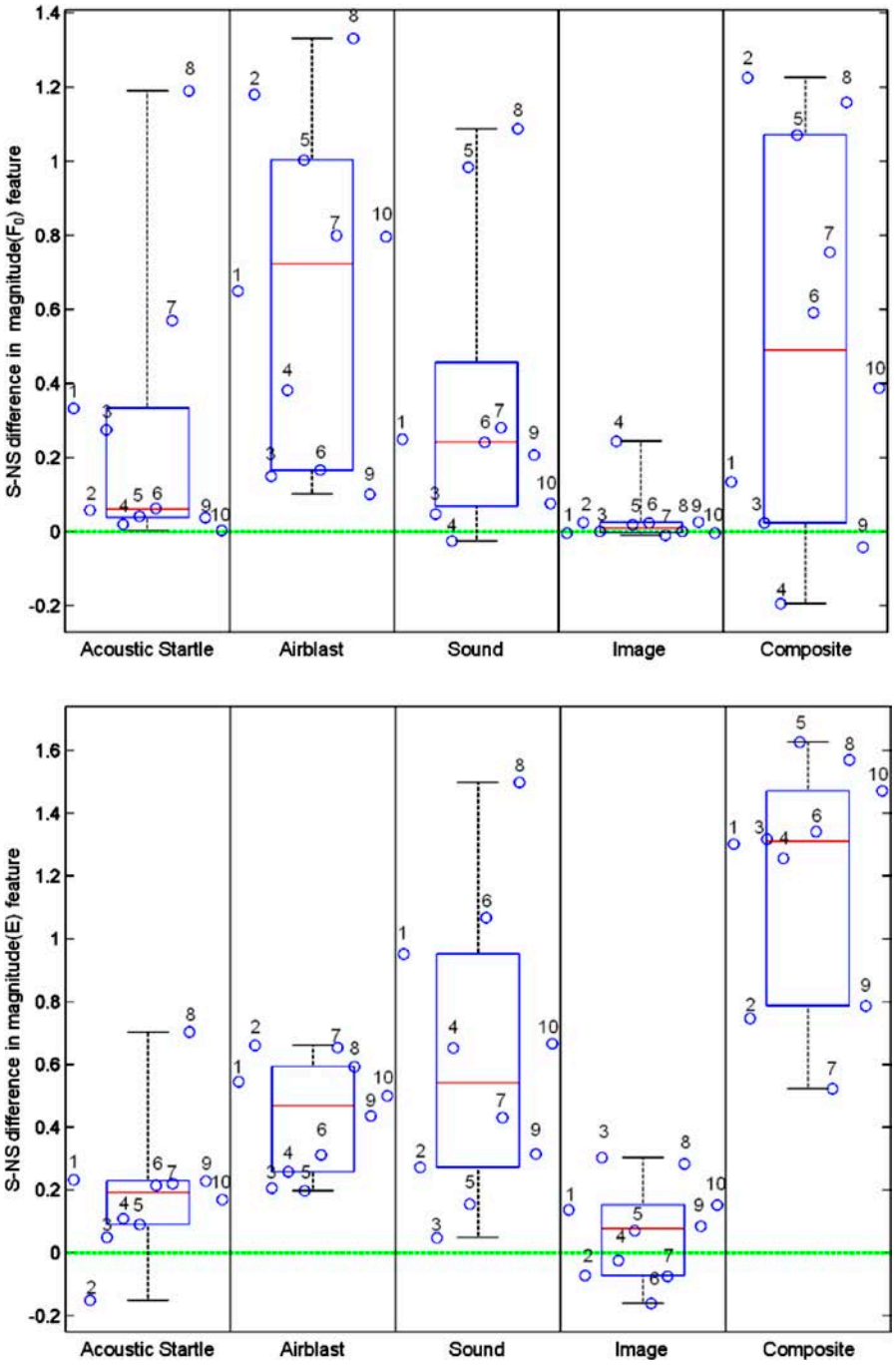


Fig. 7 – Distribution of the differences between each participant's average magnitude(F0) in Stimulus (S) and No Stimulus (NS) conditions, for each paradigm (top panel); the same for magnitude(E) (bottom panel). Numbers 1-10 indicate individual participants.

The results of the Cohen's d_z effect size measure (Equation 4) for speech features are shown in Table 2. For the magnitude(F_0), the ranking of paradigms was the same as for median difference: (1) airblast – highest; (2) composite; (3) sound; (4) acoustic startle; (5) image. On the other hand, the ranking of paradigms was slightly different than the median difference in the case of the magnitude(E): (1) composite – highest; (2) airblast; (3) sound; (4) acoustic startle; (5) image. Only the airblast and sound paradigms switched between 2nd and 3rd places in this case.

Table 2 – Effect size of startle type stimuli on magnitude(F_0) and magnitude(E) (Cohen's d_z, N = 10).

	Acoustic startle	Airblast	Sound	Image	Composite	Average
magnitude(F_0)	0.690	**1.478**	0.823	0.424	0.968	0.877
magnitude(E)	0.863	2.409	1.326	0.451	**3.171**	1.644

After applying one-tailed one-sample t-tests to the data from each paradigm in Figure 7, the greater-than-zero effects of stimuli were statistically significant ($p < 0.05$) for all paradigms except the image paradigm ($p = 0.106$ for magnitude(F_0), $p = 0.094$ for magnitude(E)). Repeated measures ANOVA rejected the null hypothesis of no difference in responding to various stimuli at 0.05 level of significance ($p = 0.012$ for magnitude(F_0), $p = 5.7*10^{-10}$ for magnitude(E)).

Furthermore, it should be noted that speech energy E had a longer time response than the F_0 (Figure 6). The reason for such different transient response lies in different phenomenology of these two speech features and the Lombard effect, i.e. the involuntary tendency of a speaker to increase the loudness, pitch, rate and duration of his/her voice when speaking in a loud noise environment [42;43].

4.3 Analysis of Eye Blinks and Head Movement

While facial video capture has been previously used in manual analysis of startle responses [44], there are only a few papers that address automated analysis of startle response videos. Bernard and colleagues proposed a novel method for automated eyelid detection and tracking in the context of eye blink startle response measurements [10]. Derakshani and Lovelace introduced their method of eyelid tracking in high-speed videos together with classification of eye blink startle responses based on support vector machines [11]. Vousdoukas and colleagues proposed stereo-vision motion tracking system for various body parts, like head, chest and shoulders, rather than individual facial features, which has been validated by comparisons with surface EMG measurements of eye blink startle response [12]. Our

approach, however, leveraged state-of-the-art general-purpose facial feature tracking technology for simultaneous tracking of facial and head movements related to startle and emotional responses.

In the current study, a low-cost off-the-shelf digital video camera, Logitech C920, was used for detection of facial components of the startle response instead of traditional EMG recordings of *orbicularis oculi* activity. While low-cost video cameras do not have time resolution of the EMG signal, video recordings of the facial expressions of the startle response can also be done with high-speed cameras (e.g. see [10]) enabling better time resolution of the *orbicularis oculi* response without the need to attach EMG electrodes to the participant's face. Therefore, in situations where the use of EMG electrodes is inconvenient, low-cost off-the-shelf digital video cameras represent a promising alternative for multimodal analyses of startle responses.

Video streams of the participants' faces acquired by the camera were processed by the state-of-the-art image processing visage|SDK FaceTrack software package (http://www.visagetechnologies.com/products/visagesdk/). The facial features tracked by visage|SDK FaceTrack engine are aligned with the Face and Body Animation part of the MPEG-4 International Standard (ISO14496). The visage|SDK FaceTrack engine also tracks head translations in x, y and z direction, rotations around x, y and z axis, eye gaze in x and y direction, as well as eye closure. Furthermore, the engine computes a variety of Action Units from the Facial Action Coding System [45], like eye blinks, lid tighteners and raisers, brow raisers and lowerers, lip stretchers, raisers and corner pullers, yaw drops, chin raisers etc. The engine has been extensively tested on the publicly available Boston database (ftp://csr.bu.edu/headtracking/), as well as on the internal test database containing 86 video sequences that amount to more than 1 hour of video [46]. The reported eye closure detection sensitivity is 88.57% and specificity is 98.26%, which is directly relevant for eye blink detection in the context of the acoustic startle reaction.

All tracked facial, head and eye features were computed in real time for each of the five stimuli paradigms from the video stream acquired by 640 x 480 resolution and frequency of 25 Hz. At the completion of each paradigm, the computed features were stored into the participants' database for subsequent off-line analysis. Among the variety of facial, head and eye features, startle type responses were the most evident in: eye blinking, lid tightening, head movement to-and-from camera, i.e. z-translation, and vertical eye gaze direction. Features that contained information regarding eye blinks and head movement were selected for more detailed analysis, because prior startle research using blank pistol shots [47] found that eye closures and head movements occurred most reliably during startle responses.

Eye blinks were measured as a binary eyeBlink time series obtained directly from the visage|SDK FaceTrack engine. It contained values 0 and 1, where 0 indicates open eye, and 1 indicates closed eye at a particular time instant. Therefore, rapid changes in eyeBlink time series from 0 to 1 and back to 0 indicated eye blinks. Head movement in the direction of the camera was obtained directly from the visage|SDK FaceTrack engine as z-translation time series. Z-translation actually indicated the distance of the participant's head from the camera, which was positioned at the top of the screen in front of the participant; therefore, larger values of z-translation indicated that the participant's head was farther away from the camera. The distance values were reported approximately in meters, as computation depends on each participant's interpupillary distance, which was not measured. Measure of the startle response in eye blinks and z-translation was defined as the range, i.e. difference between maximum and minimum value, during the interval (0 s, 0.5 s) after the onset of the startle stimulus, which is the same as the interval (3 s, 3.5 s) after the appearance of the white screen with fixation cross '+'.

Segments of eyeBlink and z-translation time series for each paradigm are shown in Figure 8, averaged across all participants in the Stimulus condition (10 participants x 3 stimuli) versus No Stimulus condition (10 participants x 3 trials without stimuli). Responses are evident immediately after the onset of startle type stimuli in all paradigms except image paradigm (e.g., see red arrows for illustration). Averaging of binary values of eyeBlink time series across all 30 trials computed the proportion of eye blink occurrences on a scale from 0 to 1.

In order to illustrate the impact of startle type stimuli, Figure 9 shows differences between each participant's eyeBlink/z-translation range when the stimuli occurred (Stimulus conditions) and the corresponding range when the stimuli did not occur (No Stimulus conditions). In each paradigm, the range was computed from the participant's average time series across 3 Stimulus conditions and, likewise, across 3 No Stimulus conditions. The largest median difference for eyeBlink range occurred in the airblast paradigm, followed by composite, acoustic startle, sound and image paradigms. Therefore, the median differences indicate that the airblast paradigm had the highest response elicitation power for eye blinks. It is also evident that, across all paradigms, participant 8 most consistently exhibited eye blink responses to startle type stimuli, while participants 4, 6 and 9 most consistently exhibited a lack of eye blink responses. The largest median difference for z-translation range occurred in the composite paradigm, followed by airblast, acoustic startle, sound and image paradigms. Therefore, the median differences indicate that the composite paradigm had the highest response elicitation power for z-translation. It is also evident that, across all paradigms, participant 8 most consistently exhibited the largest z-translation responses to startle stimuli, while participants 4, 6 and 10 most consistently exhibited the smallest responses.

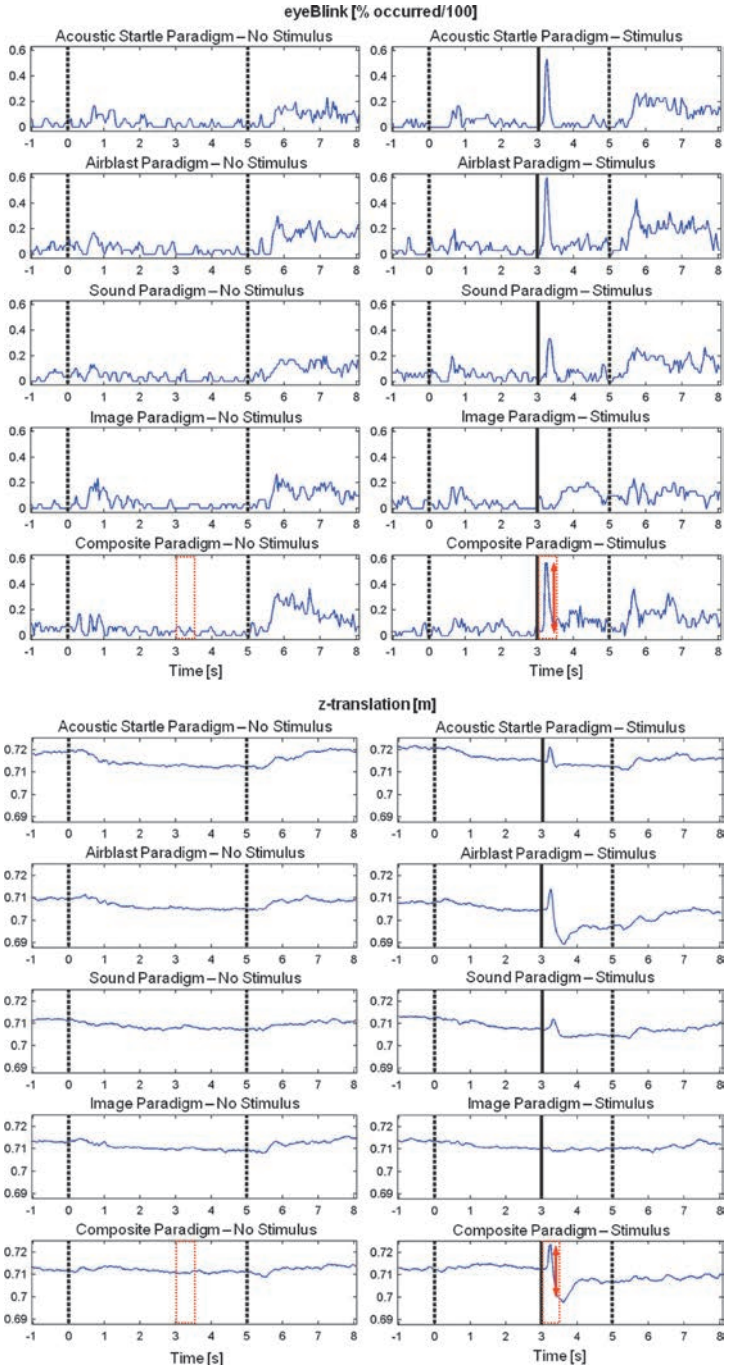


Fig. 8 – Average eyeBlink (top panel) and z-translation (bottom panel) time series across all participants before, during and after vocalizations, for trials without stimuli presented (No Stimulus columns) and for trials when stimuli occurred (Stimulus columns).

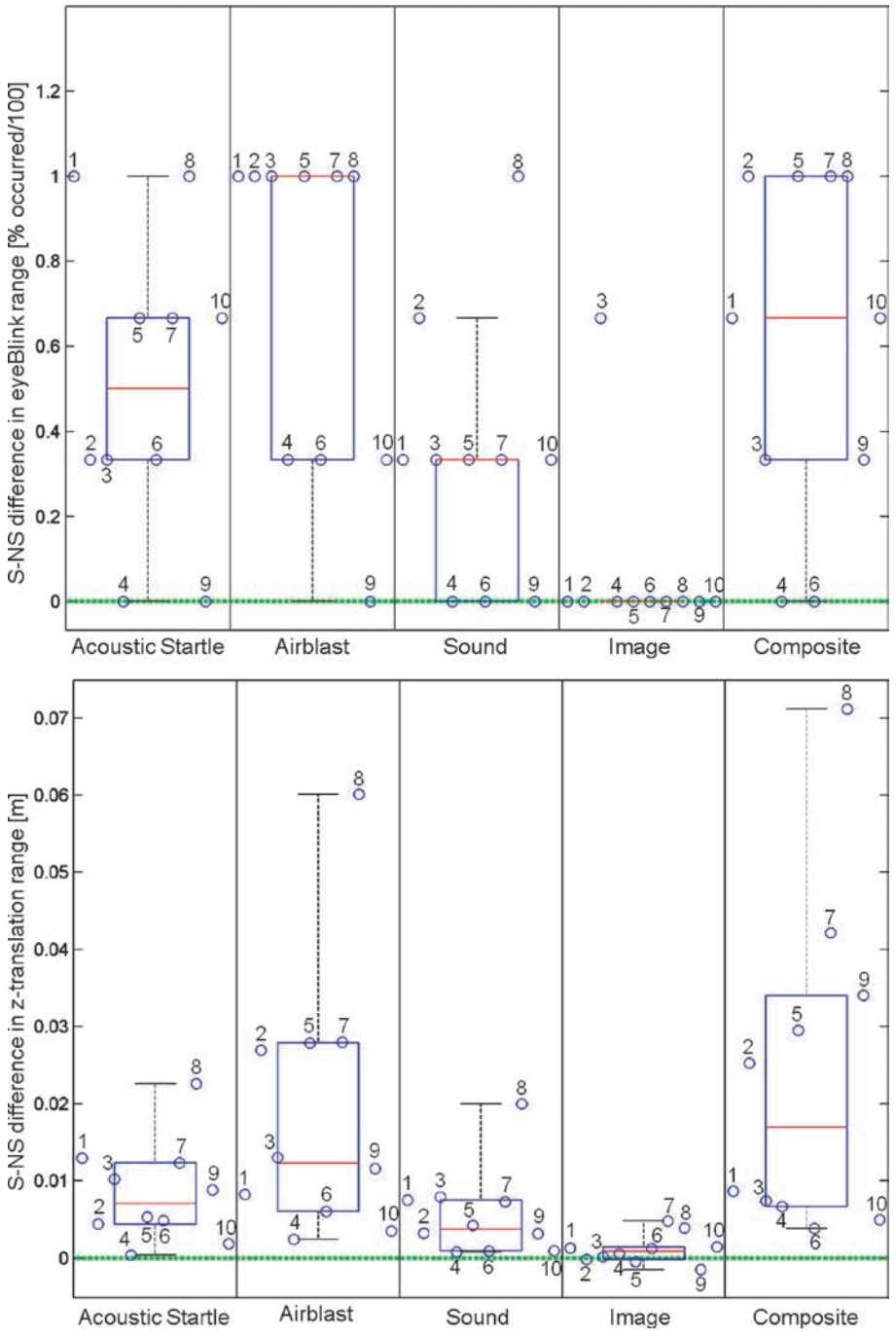


Fig. 9 – Distribution of the differences between each participant's average eyeBlink range in Stimulus (S) and No Stimulus (NS) conditions, for each paradigm (top panel); the same for z-translation range (bottom panel). Numbers 1–10 indicate individual participants.

The results of the Cohen's d_z effect size measure (Equation 4) for facial features are shown in Table 3. For eyeBlink range, the ranking of paradigms according to Cohen's d_z was the same as for median difference: (1) airblast – highest; (2) composite; (3) acoustic startle; (4) sound; (5) image. For z-translation range, the ranking of paradigms according to Cohen's d_z is: (1) acoustic startle – highest; (2) composite; (3) airblast; (4) sound; (5) image. The obtained ranking was therefore somewhat different in comparison to median difference as a measure of the response elicitation power of the paradigms.

Table 3 – Effect size of startle type stimuli on eyeBlink and z-translation range (Cohen's d_z, N = 10).

	Acoustic startle	Airblast	Sound	Image	Composite	Average
eyeBlink range	1.388	**1.754**	1.060	0.316	1.464	1.196
z-translation range	**1.275**	1.064	0.965	0.572	1.072	0.990

After applying one-tailed one-sample t-tests to the data from each paradigm in Figure 9, the greater-than-zero effects of stimuli were statistically significant ($p < 0.05$) for all paradigms except the image paradigm ($p = 0.171$ for eyeBlink range, $p = 0.052$ for z-translation range). Repeated measures ANOVA rejected the null hypothesis of no difference in responding to various stimuli at 0.05 level of significance ($p = 0.0017$ for eyeBlink range, $p = 0.0020$ for z-translation range).

5. Conclusions

The system for evaluation of startle type responses based on multimodal analyses of predominant physiological, acoustic and facial features was developed and tested on a sample of first-year female students of psychology. Subjects with higher emotional scores were selected in order to demonstrate the effectiveness of the developed system and experimental paradigms for multimodal response elicitation and analysis of startle type responses. Among physiological features, skin conductance response revealed the largest differences between Stimulus and No Stimulus conditions. The most evident Stimulus vs. No Stimulus differences in speech features were observed in fundamental frequency (F_0) and root-mean-square energy. The most significant differences in facial features occurred in eye blinks and forward-backward head movement. Among the five selected paradigms, the composite paradigm composed of airblast and congruent aversive images and sounds induced more stress in comparison with other paradigms causing higher activation of the sympathetic nervous system. At the same time, the paradigm containing only

aversive image stimuli exhibited the lowest sensitivity, i.e. it was the paradigm in which startle response was not observed at all. The largest effect size for Stimulus vs. No Stimulus conditions across all paradigms and participants was observed for speech energy, followed by eye blink, skin conductance, z-translation and F_0. It was also evident that, across all paradigms, some participants were more sensitive responders to the startle type stimuli, while other participants were less sensitive responders. Even in a relatively homogeneous group with statistically relevant similarities between individuals, there were visible individual differences, which might indicate more or less individual resiliency/vulnerability to aversive startle type stimuli. Different dynamics of specific response channels were visible from the multimodal analyses of startle type response; e.g., eye blink exhibited a short response time versus the longer response time of skin conductance.

Decreased or attenuated skin conductance startle type responses over time were an indication of decreasing sympathetic activation. However, it could be expected that different startle modulations as well as different order of presented paradigms might also have influence on overall startle type responses and response habituation. All together, it means that emotional and attentional factors may strongly interfere with reflexive response, making the overall net effect of startle response either more or less intense.

Overlaying aversive stressors on basic startle stimuli warrants more exploration in applied research regarding human resilience or vulnerability to stress. In this paper, a composite paradigm which includes congruent negatively valenced and highly arousing IAPS images/IADS sounds superimposed on airblasts showed the highest response elicitation power. These superimposed emotional contents could increase or decrease startle response, i.e. modulate physiological [48] and other multimodal variables. Therefore, enhancing power of response elicitation by semantically and contextually specific stimuli, as well as cross-correlation analysis of context and semantics of selected startle type stimuli with individually relevant context and semantics, deserves more attention in future work, particularly in broader research on cognitive regulation of emotion [49;50]. Other viable research directions may address generalizability of findings obtained in this paper, by multimodal analyses of startle type responses in males versus females or in different age groups. Likewise, these analyses can be applied in clinical research setting, e.g. with posttraumatic stress disorder (PTSD) patients and healthy controls, which is the specific focus of our ongoing research collaboration with the Department of Psychiatry, University Hospital Center Zagreb. Outlined research directions are compatible with the developed research system demonstrated in this paper, which has been designed as a general purpose test-bed for analysis of startle type responses, as well as multimodal emotion elicitation and analysis.

Acknowledgements

We are grateful to Prof. Goran Sedmak, Prof. Miloš Judaš and Academician Ivica Kostović from Croatian Institute for Brain Research for their valuable comments, as well as Sonja Grđan, M.Sc. for assistance during the laboratory experimentation. We also thank the Adris Foundation for financial support in development of the laboratory research infrastructure. Finally, our thanks go to Prof. Tanja Jovanovic, Emory University School of Medicine, for valuable comments and suggestions.

References

[1] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and Hudspeth A J., *Principles of neural science*, 5 ed. New York, NY: McGraw-Hill, 2012.

[2] M. Koch, "The neurobiology of startle," *Progress in Neurobiology*, vol. 59, no. 2, pp. 107-128, 1999.

[3] S. Holand, A. Girard, D. Laude, C. Meyer-Bisch, and J. L. Elghozi, "Effects of an auditory startle stimulus on blood pressure and heart rate in humans," *Journal of Hypertension*, vol. 17, no. 12, pp. 1893-1897, 1999.

[4] O. V. Lipp, D. A. Siddle, and P. J. Dall, "Effects of stimulus modality and task condition on blink startle modification and on electrodermal responses," *Psychophysiology*, vol. 35, no. 4, pp. 452-461, 1998.

[5] C. Herbert and J. Kissler, "Motivational priming and processing interrupt: Startle reflex modulation during shallow and deep processing of emotional words," *International Journal of Psychophysiology*, vol. 76, no. 2, pp. 64-71, 2010.

[6] M. W. Miller, C. J. Patrick, and G. K. Levenston, "Affective imagery and the startle response: Probing mechanisms of modulation during pleasant scenes, personal experiences, and discrete negative emotions," *Psychophysiology*, vol. 39, no. 4, pp. 519-529, 2002.

[7] M. Horvat, N. Bogunović, and K. Ćosić, "STIMONT: a core ontology for multimedia stimuli description," *Multimedia Tools and Applications*, vol. 73, no. 3, pp. 1103-1127, 2014.

[8] G. W. Alpers, D. Adolph, and P. Pauli, "Emotional scenes and facial expressions elicit different psychophysiological responses," *International Journal of Psychophysiology*, vol. 80, no. 3, pp. 173-181, 2011.

[9] S. Sapir, M. D. McClean, and C. R. Larson, "Human laryngeal responses to auditory stimulation," *The Journal of the Acoustical Society of America*, vol. 73, no. 1, pp. 315-321, 1983.

[10] F. Bernard, C. E. Deuter, P. Gemmar, and H. Schachinger, "Eyelid contour detection and tracking for startle research related eye-blink measurements from high-speed video records," *Computer Methods and Programs in Biomedicine*, vol. 112, no. 1, pp. 22-37, 2013.

[11] R. R. Derakhshani and C. T. Lovelace, "An ensemble method for classifying startle eyeblink modulation from high-speed video records," *IEEE Transaction on Affective Computing*, vol. 2, no. 1, pp. 50-63, 2011.

[12] M. I. Vousdoukas, P. Perakakis, S. Idrissi, and J. Vila, "SVMT: A MATLAB toolbox for stereo-vision motion tracking of motor reactivity," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 318-329, 2012.

[13] T. Jovanovic, S. D. Norrholm, N. Q. Blanding, M. Davis, E. Duncan, B. Bradley, and K. J. Ressler, "Impaired fear inhibition is a biomarker of PTSD but not depression," *Depression and Anxiety*, vol. 27, no. 3, pp. 244-251, 2010.

[14] S. D. Norrholm, B. Vervliet, T. Jovanovic, W. Boshoven, K. M. Myers, M. Davis, B. Rothbaum, and E. J. Duncan, "Timing of extinction relative to acquisition: a parametric analysis of fear extinction in humans," *Behavioral Neuroscience*, vol. 122, no. 5, pp. 1016-1030, 2008.

[15] M. Horvat, "Generation of multimedia stimuli based on ontological affective and semantic annotation." Doctoral thesis, University of Zagreb, 2013.

[16] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual." Technical Report A-8, University of Florida, Gainesville, FL, 2008.

[17] M. M. Bradley and P. J. Lang, "The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual." Technical Report B-3. University of Florida, Gainesville, FL,2007.

[18] N. Tottenham, J. W. Tanaka, A. C. Leon, T. McCarry, M. Nurse, T. A. Hare, D. J. Marcus, A. Westerlund, B. J. Casey, and C. Nelson, "The NimStim set of facial expressions: judgments from untrained research participants," *Psychiatry Research*, vol. 168, no. 3, pp. 242-249, 2009.

[19] A. Marchewka, Ł. Żurawski, K. Jednoróg, and A. Grabowska, "The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database," *Behavior Research Methods*, vol. 46, no. 2, pp. 596-610, 2014.

[20] D. F. Ramirez-Moreno and T. J. Sejnowski, "A computational model for the modulation of the prepulse inhibition of the acoustic startle reflex," *Biological Cybernetics*, vol. 106, no. 3, pp. 169-176, 2012.

[21] S. D. Norrholm, T. Jovanovic, I. W. Olin, L. A. Sands, B. Bradley, and K. J. Ressler, "Fear extinction in traumatized civilians with posttraumatic stress disorder: relation to symptom severity," *Biological Psychiatry*, vol. 69, no. 6, pp. 556-563, 2011.

[22] K. C. Wilkins, A. J. Lang, and S. B. Norman, "Synthesis of the psychometric properties of the PTSD Checklist (PCL) military, civilian, and specific versions," *Depression and Anxiety*, vol. 28, no. 7, pp. 596-606, 2011.

[23] A. M. Kring, D. A. Smith, and J. M. Neale, "Individual differences in dispositional expressiveness: Development and validation of the Emotional Expressivity Scale," *Journal of Personality and Social Psychology*, vol. 66, no. 5, pp. 934-949, 1994.

[24] T. Jurin, N. Jokic-Begic, and A. L. Korajlija, "Factor structure and psychometric properties of the anxiety sensitivity index in a sample of Croatian adults," *Assessment*, vol. 19, no. 1, pp. 31-41, 2012.

[25] S. Reiss, R. A. Peterson, D. M. Gursky, and R. J. McNally, "Anxiety sensitivity, anxiety frequency and the prediction of fearfulness," *Behaviour Research and Therapy*, vol. 24, no. 1, pp. 1-8, 1986.

[26] P. K. Pilz, T. D. Carl, and C. F. Plappert, "Habituation of the acoustic and the tactile startle responses in mice: two independent sensory processes," *Behavioral Neuroscience*, vol. 118, no. 5, p. 975, 2004.

[27] B. Valsamis and S. Schmid, "Habituation and prepulse inhibition of acoustic startle in rodents," *Journal of Visualized Experiments: JoVE*, no. 55, p. e3446, 2011.

[28] J. T. Cacioppo and L. G. Tassinary, "Inferring psychological significance from physiological signals," *American Psychologist*, vol. 45, no. 1, pp. 16-28, 1990.

[29] P. J. Lang, "The emotion probe: Studies of motivation and attention," *American Psychologist*, vol. 50, no. 5, pp. 372-385, 1995.

[30] D. Kukolja, S. Popović, M. Horvat, B. Kovač, and K. Ćosić, "Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications," *International Journal of Human-Computer Studies*, vol. 72, no. 10, pp. 717-727, 2014.

[31] W. Boucsein, *Electrodermal activity*, 2 ed. Boston, MA: Springer Verlag, 2011.

[32] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs," *Frontiers in Psychology*, vol. 4, no. 863,Published online 2013 November 26. doi: 10.3389/fpsyg.2013.00863

[33] T. Baer, "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes," *The Journal of the Acoustical Society of America*, vol. 65, no. 5, pp. 1271-1275, 1979.

[34] J. H. Martin, J. E. Thomson, J. E. Aviv, T. Kim, B. Diamond, R. L. Sacco, and L. G. Close, "Laryngopharyngeal sensory discrimination testing and the laryngeal adductor reflex," *Annals of Otology, Rhinology & Laryngology*, vol. 108, no. 8, pp. 725-730, 1999.

[35] C. L. Ludlow, F. Van Pelt, and J. Koda, "Characteristics of late responses to superior laryngeal nerve stimulation in humans," *Annals of Otology, Rhinology & Laryngology*, vol. 101, no. 2, pp. 127-134, 1992.

[36] T. D. Blumenthal, B. N. Cuthbert, D. L. Filion, S. Hackley, O. V. Lipp, and A. Van Boxtel, "Committee report: Guidelines for human startle eyeblink electromyographic studies," *Psychophysiology*, vol. 42, no. 1, pp. 1-15, 2005.

[37] K. M. Berg, "Elicitation of acoustic startle in the human." Doctoral thesis, University of Wisconsin, 1973.

[38] T. D. Blumenthal, "The startle response to acoustic stimuli near startle threshold: effects of stimulus rise and fall time, duration, and intensity," *Psychophysiology*, vol. 25, no. 5, pp. 607-611, 1988.

[39] B. Dropuljić, "Emotional state estimation based on data mining of acoustic speech features." Doctoral thesis, University of Zagreb, 2014.

[40] B. Dropuljić, I. Mijić, D. Petrinović, K. Ćosić, and T. Jovanovic, "Comparative analysis of orbicularis oculi electromyogram and voice fundamental frequency variation in the context of acoustic startle response," 5th IEEE Conference on Cognitive Infocommunications, Vietri sul Mare, Italy, November 5-7, 2014, pp. 149-154.

[41] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," Proceedings of the 21st ACM international conference on Multimedia, Barcelona, Spain, 2013, pp. 835-838.

[42] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," *Journal of Speech, Language, and Hearing Research*, vol. 14, no. 4, pp. 677-709, 1971.

[43] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," *The Journal of the Acoustical Society of America*, vol. 84, no. 3, pp. 917-928, 1988.

[44] J. Blechert, T. Michael, and F. H. Wilhelm, "Video-based analysis of bodily startle and subsequent emotional facial expression in posttraumatic stress disorder," *Journal of Experimental Psychopathology*, vol. 4, no. 4, pp. 435-447, 2013.

[45] P. Ekman and W. V. Freisen, *Facial action coding system: a technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[46] Visage Technologies Face Tracking and Animation, "Face Tracking accuracy and performance overview: visage|SDK v7.2." July 18, 2014.

[47] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*, 2 ed. New York, NY: Oxford University Press, 2005.

[48] A. Blanch, F. Balada, and A. Aluja, "Presentation and AcqKnowledge: An application of software to study human emotions and individual differences," *Computer Methods and Programs in Biomedicine*, vol. 110, no. 1, pp. 89-98, 2013.

[49] K. N. Ochsner and J. J. Gross, "Cognitive emotion regulation: Insights from social cognitive and affective neuroscience," *Current Directions in Psychological Science*, vol. 17, no. 2, pp. 153-158, 2008.

[50] K. Ćosić, S. Popović, B. Kovač, D. Kukolja, D. Ivanec, and T. Jovanovic, "System for evaluation of cognitive performance under the emotional stressors," COGNITIVE 2014: The Sixth International Conference on Advanced Cognitive Technologies and Applications, Venice, Italy, May 25-29, 2014, pp. 239-245.

Activities of the Croatian Academy of Engineering in 2015

Meetings of the Academy's Bodies,
Academy's Auspices, Organization and Co-Organization of Meetings,
and Major Meetings of Public Interest in 2015

Meetings of the Academy's Bodies

The majority of the Academy's activities takes place in the Academy's Bodies: Departments, Standing Committees and Centers of the Academy, the Governing Board of the Academy, the Presidency of the Academy, the Scientific Council of the Academy and the Assembly of the Academy.

In 2015 the Departments, Standing Committees and Centers have held more than 30 meetings in total. The activities included especially the procedures of election of new Emeriti, Members, Associates, International and Supporting Members of the Academy as well as procedures of evaluation of candidates for the Awards of the Academy and preparation of papers for the "Annual 2015 of the Croatian Academy of Engineering". The activities have also been strongly focused on the international cooperation of the Academy. Several Departments, Committees and Centers have prepared materials for the Guest Editorial participation in the Academy's Bulletin "Engineering Power".

The Governing Board has had 20 meetings, and the Presidency of the Academy has held 4 meetings, respectively, dealing with the most imminent as well as vital and long-term organizational and decision-making matters of the Academy's scientific, expert and professional activities, membership, finances etc.

The 30th Annual (Elective) Assembly of the Academy has been held on May 13, 2015 in the Great Hall of the University of Zagreb.

The Scientific Council of the Academy has had 2 meetings in 2015, dealing with domestic and international scientific cooperation and domestic scientific institutions' status and evaluation procedure. The Council had proposed peer-reviewers

for the Academy's Awards candidacies and nominated members from the Academy's Departments to the editorial bodies of the Croatian Technological Encyclopedia, which is the joint project of the Croatian Academy of Sciences and Arts, Miroslav Krleža Institute of Lexicography and Croatian Academy of Engineering.

Academy's Auspices

- Faculty of Textile Technology in Zagreb – Scientific and Professional Colloquium "Textile Science and the Economy: Functional Materials, Clothing, Footwear and Accessories" (Zagreb, January 26, 2015)
- Faculty of Mechanical Engineering and Naval Architecture in Zagreb – Final Conference of the Projekt "Additive Technologies for Small and Medium Enterprises – AdTecSME" (Zagreb, March 11, 2015)
- 5th Domestic and 1st International Scientific and Professional Meeting "Water for All" (Osijek, March 20, 2015)
- Croatian Colours Association - International Colours Day 2015 (Zagreb, March 21, 2015)
- Croatian Association of Chemical Engineers and Technologists – 25th Croatian Meeting of Chemists and Chemical Engineers (Zagreb, Faculty of Chemical Engineering and Technology, April 21-24, 2015)
- 13th European Transport Congress (Zagreb, April 23-24, 2015)
- Croatian Society of Cartography – 11th Colloquium on Cartography and Geoinformation (Buzet, May 8, 2015)
- Croatian Energy Society – Presentation of Hrvoje Požar Foundation Annual Awards (Zagreb, July 6, 2015)
- Faculty of Geodesy in Zagreb – Professional and Scientific Meeting "Shareholders' Profile in Geodesy 3: Technological Development and Professions in Geodesy and Geoinformatics" (Zagreb, October 2, 2015)
- Faculty of Civil Engineering in Zagreb and Technological University in Prague – International Conference "Applications of Structural Fire Engineering – ASFE 2015" (Dubrovnik, October 15-16, 2015
- Faculty of Forestry in Zagreb – 26th International Conference on Wood Science and Technology "Implementation of Wood Science in Woodworking Sector" (within "Ambienta '15", Zagreb, Zagreb Fair, October 16, 2015)

- Faculty of Food Technology in Osijek and ICC – 8th International Congress "Flour-Bread '15" and 10th Croatian Congress of Flour Production and Processing Technologists "Flour-Bread '15" (Opatija, October 28-30, 2015)
- Faculty of Geodesy in Zagreb – 2nd Interdisciplinary Scientific Conference "The Western Balkans Meets the EU: Ongoings inside Geospatial Domain and Sustainable Development" (Zagreb, November 26-27, 2015)

Organization and Co-Organization of Meetings

- HATZ Department of Chemical Engineering and Pliva Croatia, Ltd., Zagreb – Scientific and Professional Meeting on Industrial Crystallization (Zagreb, January 23, 2016)
- HATZ Department of Systems and Cybernetics – Forum "Intelligent Transportation Systems" by Prof. Sadko Mandžuka, Ph. D. (Zagreb, February, 2016)
- 1st Joint Session of the Council and Coordination of the Four Academies (Croatian Academy of Engineering – HATZ, Academy of Medical Sciences of Croatia – AMZH, Academy of Legal Sciences in Croatia – APZH and Academy of Forestry Sciences – AŠZ) (HATZ, Zagreb, February 5th, 2015)
- HATZ and Croatian Engineers' Association - The Republic of Croatia Engineers' Day 2015 (under the auspices of HAZU and the Croatian Ministry of Economy) (Zagreb, March 2, 2015)
- HATZ Department of Transportation, HATZ Center for Traffic Engineering and Croatian Railway Traffic Employees' Syndicate – Round Table Discussion "The Necessity of Construction of the Lowland, Two-Gauge and Electrified Railway Rijeka-Zagreb-Botovo as a Section of the Corridor 11: Baltic-Adriatic i.e. Route C-65" (Rijeka, March 23, 2015)
- HATZ Department of Systems and Cybernetics – Forum "ICT in the Economy and at the University" by Prof. Željko Hocenski, Ph. D., Prof. Franjo Jović, Ph. D. and Prof. Niko Majdandžić, Ph. D.
- HATZ Department of Graphical Engineering and HATZ Center for Graphical Engineering – Symposium "Printing & Design 2015" (Zagreb, March 27, 2015)
- HATZ Department of Systems and Cybernetics and Faculty of Electrical Engineering and Computing in Zagreb – Forum "Elecrical Brain Signals" by Prof. Mario Cifrek, Ph. D. (Zagreb, April, 2015)
- 2nd Joint Session of the Council and Coordination of the Four Academies (HATZ, AMZH, APZH and AŠZ) (HATZ, Zagreb, April 13, 2015)

- Coorganization by Four Academies – Forum "Multidisciplinary Metrics for Prediction of Mental Resilience for Performing Stressful Jobs" (Zagreb, April 14, 2015)
- HATZ Department of Graphical Engineering, Graphical Faculty in Zagreb and Polytechnics in Zagreb – Lecture by Prof. Rajendrakumar Anayath, Ph. D., "A Foray into the Concept of Quality and How Its Focus Changed from Time to Time" (Zagreb, June 2, 2015)
- HAZU Scientific Council for Technological Development and HATZ Department of Systems and Cybernetics – Lecture "Cognitive Machines – A Challenge of Trans-Disciplinarity" by Prof. Bojan Jerbić, Ph. D. (June 11, 2015)
- Joint Forum by the Council of the Four Academies (APZH, HATZ, AMZH i AŠZ) - Prof. Davor Derenčinović, Ph. D., "Croatian Legal System After the Accession of the Republic of Croatia to the European Union" (Zagreb, June 16, 2015)
- 3rd Joint Session of the Council and Coordination of the Four Academies (HATZ, AMZH, APZH and AŠZ) (HATZ, Zagreb, June 29, 2015)
- HATZ Committee for Cooperation with the Economy and Regional Cooperation and Croatian Chamber of Commerce – Round Table Discussion "Application of New Technology in the Development of Alimentary Products – Situation in the EU and the Perspectives in the Republic of Croatia by 2020) (Zagreb, July 1, 2015)
- HATZ and HAZU – Round Table Discussion on Life and Work of Faust Vrančić (Zagreb, November 5, 2015)
- HATZ Department of Systems and Cybernetics and Faculty of Electrical Engineering and Computing in Zagreb – Presentation of the Book "Geometrical Linear Holography" written by Prof. Franjo Jović, Ph. D. (Zagreb, November 23, 2015)

Participation at the Major Meetings of Public Interest

- Croatian Academy of Sciences and Arts (HAZU) – Mini-Symposium "Science in Croatia" (Zagreb, February 5, 2015)
- Office of the President of the Republic of Croatia - Ceremony of Decorations Awarding (Zagreb, February 9, 2015)
- HAZU and Ruđer Bošković Institute – Commemorative Meeting for the Late Academician Branko Souček (Zagreb, February 10, 2015)

- Faculty of Civil Engineering in Zagreb - Jubilary Session anent the Faculty Day (Zagreb, February 20, 2015)
- Academy of Medical Sciences of Croatia (AMZH) – Forum "Brain and the Art" by Academician Vida Demarin (Zagreb, February 24, 2015)
- HAZU – Round Table Discussion "National Research and Innovation Infrastructure in the Strategy of Education, Science and Technology" (Zagreb, February 26, 2015)
- Faculty of Electrical Engineering and Computing (FER) - Public Presentation of the Projects of the Faculty of Electrical Engineering and Computing in Zagreb, Co-Funded by the European Fund for Regional Development within the EFRD Call "Strengthening the Capacities for Research, Development and Innovation" (Zagreb, February 26, 2015)
- EURAXESS – EURAXESS RISE Information Day (Zagreb, March 6, 2015)
- HAZU – G. I. D.-EMAN Conference Parmenides VII – Common Heritage and Technologies: Enhancement of Heritage, a Key to Development (Dubrovnik, March 16-19, 2015)
- HAZU – Lecture by Academician Vida Demarin "A Brain That Lasts – The Newest Understandings on Brain Health Preservation" (Zagreb, March 19, 2015)
- Forum "DNA Analysis in the Function of Convicts' Rights Protection: American Experiences and Croatian Perspectives" (Zagreb, March 20, 2015)
- Faculty of Textile Technology in Zagreb (TTF) – Scientific Forum "The New Classification of Substances and Materials – Polymers and Neo-Polymers" (Zagreb, March 26, 2015)
- HAZU Department for Medical Sciences and AMZH – Forum "Croatian Stomatology since Austro-Hungarian Monarchy till the European Union" (Zagreb, March 31, 2015)
- Faculty of Science in Zagreb - Department of Geology and Institute of Geology and Paleontology – International Conference anent the 100th Anniversary of Birth of Academician Vanda Kochansky-Devidé (Zagreb, April 9, 2015)
- Croatian Cartographic Society (CCS) and HATZ – Session of the CCS Presidency (Zagreb, April 10, 2015)
- Meeting of the Representatives of the Four Academies (Croatian Academy of Engineering, Academy of Medical Sciences of Croatia, Academy of Legal Sciences of Croatia and Academy of Forestry Sciences) with the Representatives of the Croatian President's Protocol (Zagreb, April 14, 2015)
- HAZU – Professional Discussion "Academic Community and the EU Funds" (Zagreb, April 15, 2015)

- Ruđer Bošković Institute – CroArtScia2015 – Technological Innovation: Art&Science" (Zagreb, May 27-30, 2015 and Sisak, May 29, 2015)
- Faculty of Civil Engineering in Zagreb – Promotion of the University Textbook written by Prof. Dubravka Bjegović, Ph. D. and Prof. Nina Štirmer, Ph. D., "Theory and Technology of Concrete" (Zagreb, June 10, 2015)
- Anniversary Celebration of the Croatian-Korean Business Club (Zagreb, June 15, 2015)
- Celebration of the 21th Energy Institute Hrvoje Požar Day (Zagreb, July 3, 2015)
- Croatian Statehood Foundation (ZHDZ) – International Security Conference 2015 (Zagreb, September 18, 2015)
- Ministry of Science, Education and Sports of the Republic of Croatia (MZOS) and University of Zagreb – Opening of the BIOCenter – Bio-Sciences and Commercialization Incubation Center (Zagreb, September 25, 2015)
- TTF – Scientific Forum "3D Printers and Their Application in Clothing Production" by Prof. Darko Gojanović, Ph. D. (Zagreb, September 25, 2015)
- Faculty of Geodesy in Zagreb – Jubilary Session anent the Faculty Day (Zagreb, September 25, 2015)
- HAZU – Lecture "Light in Graphical Technology and Visual Arts" by Prof. Emer. Vilko Žiljak, Ph. D. (within the Symposium "Man and Light") (Zagreb, September 29, 2015)
- Faculty of Transportation and Traffic Sciences – The Faculty Day (Zagreb, October 12, 2015)
- CAETS Council Meeting (New Delhi, India, October 12-16, 2015)
- Jubilary Session anent the 96th Anniversary of the Faculty of Chemical Engineering and Technology in Zagreb (Zagreb, October 20, 2015)
- Croatian Gas Association and INA, Inc. - Working and Jubilary Session anent 13th Gas Days (Zagreb, October 20, 2015)
- Faculty of Forestry in Zagreb – Jubilary Session anent the 117th Anniversary and the Faculty Day (Zagreb, October 23, 2015)
- HAZU – Celebration of the 200th Anniversary of Birth of Bishop Josip Juraj Strossmayer (Zagreb, October 27, 2015)
- University of Zagreb – Dies Academicus / University of Zagreb Day (Zagreb, October 30, 2015)
- Euro-CASE Annual Conference „Engineering Smart Cities of the Future", Board Meeting and Executive Committee Meeting (Delft and The Hague, Netherlands, November 2-3, 2015)

- MZOS – National Technological Culture Award "Faust Vrančić" 2014 Presentation Ceremony (Zagreb, November 9, 2015)
- HAZU Scientific Council for Technological Development Elective Assembly and Lecture by Prof. Nedjeljko Perić, Ph. D., "Nikola Tesla Innovation Center – A Link Connecting Science and the Economy" (Zagreb, November 17, 2015)
- Faculty of Electrical Engineering and Computing in Zagreb – The Faculty Day Celebration (Zagreb, November 20, 2015)
- AMZH – Forum "Onco-Plastic Approach to Surgical Treatment of Breast Cancer" by Prof. Zdenko Stanec, Ph. D. (Zagreb, November 24, 2015)
- Faculty of Geodesy in Zagreb - 2nd International Interdisciplinary Scientific Conference "Western Balkans Meets the EU: Ongoings Inside Geodetic Domain and Sustainable Development" (Zagreb, November 26-27, 2015)
- Faculty of Mining, Geology and Petroleum Engineering in Zagreb (RGNF) – The Faculty Day (Zagreb, December 4, 2015)