UNIVERSITY OF WARMIA AND MAZURY IN OLSZTYN

Technical Sciences

21(1) 2018

PUBLISHER UWM

OLSZTYN 2018

The Technical Sciences is indexed and abstracted in BazTech (http://baztech.icm.edu.pl) and in IC Journal Master List (http://journals.indexcopernicus.com)

The Journal is available in electronic form on the web sites
http://www.uwm.edu.pl/techsci (subpage Issues)
http://wydawnictwo.uwm.edu.pl (subpage Czytelnia)

The electronic edition is the primary version of the Journal

PL ISSN 1505-4675
e-ISNN 2083-4527

Address
ul. Jana Heweliusza 14
10-718 Olsztyn-Kortowo, Poland
tel.: +48 89 523 36 61
fax: +48 89 523 34 38
e-mail: wydawca@uwm.edu.pl

Ark. wyd. 5,9, ark. druk. 4,75, nakład 85 egz.
Druk – Zakład Poligraficzny UWM w Olsztynie
zam. nr 89

Contents

Technical Sciences, 2018, 21(1), 5–17

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Quarterly peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: www.uwm.edu.pl/techsci/

MONTE CARLO SIMULATION OF CLIMATE-WEATHER CHANGE PROCESS AT MARITIME FERRY OPERATING AREA

Ewa Kuligowska

Department of Mathematics
Gdynia Maritime University

Received 17 March 2016, accepted 25 October 2017, available online 6 November 2017.

Key words: Monte Carlo simulation, climate-weather change process, critical infrastructure operating area.

Abstract

The paper presents a computer simulation technique applied to generating the climate-weather change process at Baltic Sea restricted waters and its characteristics evaluation. The Monte Carlo method is used under the assumption of semi-Markov model of this process. A procedure and an algorithm of climate-weather change process' realizations generating and its characteristics evaluation are proposed to be applied in C# program preparation. Using this program, the climate-weather change process' characteristics are predicted for the maritime ferry operating area. Namely, the mean values and standard deviations of the unconditional sojourn times, the limit values of transient probabilities and the mean values of total sojourn times for the fixed time at the climate-weather states are determined.

Symbols:

$C(t)$ – climate weather change process,
c_b – climate-weather state,
w – number of climate-weather states,
Ξ_{bl} – random conditional sojourn times of a process $C(t)$ at climate-weather states c_b, when its next state is c_l,
$\xi_{bl}^{(k)}$ – realization of the conditional sojourn time Ξ_{bl}, of a process $C(t)$,
ξ – experiment time,
n_{bl} – number of sojourn time realizations during the time ξ,
$[C_{bl}(t)]_{w\times w}$ – matrix of conditional distribution functions of conditional sojourn times Ξ_{bl},
$c_{bl}(t)$ – conditional density function of the distribution function $C_{bl}(t)$,
$C_{bl}^{-1}(h)$ – inverse function of the distribution function $C_{bl}(t)$,

Correspondence: Ewa Kuligowska, Katedra Matematyki, Akademia Morska, ul. Morska 83, 81-225 Gdynia, phone: +48 58 55 86 232, e-mail: e.kuligowska@wn.am.gdynia.pl

g, h, h_1, h_2 – randomly generated numbers from the interval $\langle 0,1)$,
Ξ_b – unconditional sojourn time of a process $C(t)$ at climate-weather state c_b,
$[q_b(0)]_{1\times w}$ – vector of probabilities of a process $C(t)$ at initial states c_b,
$[q_{bl}]_{w\times w}$ – matrix of probabilities of transitions of a process $C(t)$ between climate-weather states c_b and c_l,
$q_b(t)$ – transient probability of a process $C(t)$ at a climate-weather state c_b at the moment t,
q_b – limit value of a transient probability $q_b(t)$,
M_b – mean value of unconditional sojourn time Ξ_b at climate-weather state c_b,
D_b – standard deviation of unconditional sojourn time Ξ_b at climate-weather state c_b,
$\hat{\Xi}_b$ – total sojourn time at climate-weather state c_b, during the fixed time,
$\hat{\xi}_b$ – realisation of the total sojourn time at climate-weather state c_b during the fixed time,
$\hat{M}_b$ – mean value of total sojourn time $\hat{\Xi}_b$ at climate-weather state cb during the fixed time.

Introduction

The general model of the climate-weather change process is proposed in (KOŁOWROCKI, SOSZYŃSKA-BUDNY 2016a). This process is defined by the initial probabilities at its states, the probabilities of transitions between these states and the distributions of the conditional sojourn times at these states. Further, the main characteristics of the considered process, i.e. the mean values and standard deviations of the unconditional sojourn times, the limit values of transient probabilities and the unconditional mean values of total sojourn times at the particular states for the fixed time can be determined. However, very often the analytical approach to the climate-weather change process' characteristics evaluation leads to complicated calculations, obtaining approximate results only (GRABSKI 2014, GRABSKI, JAŹWIŃSKI 2009, KOŁOWROCKI, KULIGOWSKA 2013, LIMNIOS, OPRISAN 2005). This paper proposes another non-analytical approximate approach, i.e. a computer simulation technique based on Monte Carlo method. This method can provide fairly accurate results in a relatively short time spent for calculations (KOŁOWROCKI et al. 2013, KROESE et al. 2011, MARSAGLIA, TSANG 2000, ZIO, MARSEGUERRA 2002). Moreover, the Monte Carlo simulation approach may be successfully applied in joint investigation of the climate-weather change process and its impact on safety of a very wide class of real critical infrastructures (KULIGOWSKA, TORBICKI 2017). To give an example of Monte Carlo simulation application, the climate-weather change process' analysis, identification and prediction at the maritime ferry operating area is performed in this paper.

Materials and methods

Climate-weather change process

We assume that the climate-weather change process for the critical infrastructure operating area is taking w, $w \in \mathrm{N}$, different climate-weather states c_1, c_2, ..., c_w. Further, we define the climate-weather change process $C(t)$, $t \in \langle 0,\infty)$, with discrete climate-weather states from the set $\{c_1, c_2, ..., c_w\}$. We assume a semi-Markov model (GRABSKI 2014, KOŁOWROCKI 2004, 2014, KOŁOWROCKI et al. 2013, KOŁOWROCKI, KULIGOWSKA 2013, KOŁOWROCKI, SOSZYŃSKA-BUDNY 2011, LIMNIOS, OPRISAN 2005), of the climate-weather change process $C(t)$ and we mark by Ξ_{bl} its conditional sojourn times at the climate-weather states c_b, when its next climate-weather state is c_l, $b, l = 1, 2, ..., w$, $b \neq l$. Under these assumptions, the climate-weather change process may be described by the following parameters:

– the vector $[q_b(0)]_{1\times w}$ of the initial probabilities $q_b(0) = P(C(0) = c_b)$, $b = 1, 2, ..., w$, of the climate-weather change process $C(t)$ staying at particular climate-weather states at the moment $t = 0$;

– the matrix $[q_{bl}]_{w\times w}$ of the probabilities q_{bl}, $b, l = 1, 2, ..., w$, $b \neq \mathrm{l}$, of the climate-weather change process $C(t)$ transitions between the climate-weather states c_b and c_l, $b, l = 1, 2, ..., w$, $b \neq l$, where by a formal agreement $q_{bb} = 0$ for $b = 1, 2, ..., w$;

– the matrix $[C_{bl}(t)]_{w\times w}$ of conditional distribution functions $C_{bl}(t) = P(\Xi_{bl} < t)$, $b, l = 1, 2, ..., w$, $b \neq l$, of the climate-weather change process $C(t)$ conditional sojourn times Ξ_{bl} at the climate-weather states, where by a formal agreement $\Xi_{bb}(t) = 0$ for $b = 1, 2, ..., w$.

Moreover, we introduce the matrix $[c_{bl}(t)]_{w\times w}$ of the density functions $c_{bl}(t)$, $b, l = 1, 2, ..., w$, $b \neq l$, of the climate-weather change process $C(t)$ conditional sojourn times Ξ_{bl}, $b, l = 1, 2, ..., w$, $b \neq \mathrm{l}$, at the climate-weather states, corresponding to the conditional distribution functions $C_{bl}(t)$.

Having in disposal the above parameters, it is possible to obtain the main characteristics of climate weather change process. From the formula for total probability, it follows that the unconditional distribution functions $C_b(t)$ of the climate-weather change process' $C(t)$ sojourn times Ξ_b, $b = 1, 2, ..., w$, at the climate-weather states c_b, $b = 1, 2, ..., w$, are given by $C_b(t) = P(\Xi_b \leq \mathrm{t}) = \sum_{l=1}^{w} q_{bl} C_{bl}(t)$, $t \in \langle 0,\infty)$, $b = 1, 2, ..., w$ (KOŁOWROCKI, SOSZYŃSKA-BUDNY 2016a). Hence, the mean values $M_b = E[\Xi_b]$ of the climate-weather change process' $C(t)$ unconditional sojourn times Ξ_b, $b = 1, 2, ..., w$, at the particular climate-weather states can be obtained (KOŁOWROCKI, SOSZYŃSKA-BUDNY 2016b). Further, the limit values of the climate-weather change process'

transient probabilities $q_b(t) = P(C(t) = c_b)$, $b = 1, 2, ..., w$, at the particular climate-weather states

$$q_b = \lim_{t\to\infty} q_b(t),\ b = 1, 2, ..., w \tag{1}$$

can be determined (KOŁOWROCKI, SOSZYŃSKA 2011).

Monte Carlo simulation approach to climate-weather change process' modelling

We denote by $c_b = c_b(g)$, $b \in \{1, 2, ..., w\}$, the realization of the climate-weather change process' initial climate-weather state at the moment $t = 0$. Further, we select this initial state by generating realizations from the distribution defined by the vector $[q_b(0)]_{1\times w}$, according to the formula

$$c_b(g) = c_i,\ \sum_{j=1}^{i} q_{j-1}(0) \le g < \sum_{j=1}^{i} q_j(0),\ i \in \{1, 2, ..., w\} \tag{2}$$

where g is a randomly generated number from the uniform distribution on the interval $\langle 0,1)$ and $q_0(0) = 0$.

After selecting the initial climate-weather state c_b, $b \in \{1, 2, ..., w\}$, we can fix the next climate-weather state of the climate-weather change process. We denote by $c_l = c_l(g)$, $l \in \{1, 2, ..., w\}$, $l \neq b$, the sequence of the realizations of the climate-weather change process' consecutive climate-weather states generated from the distribution defined by the matrix $[q_{bl}]_{w\times w}$. Those realizations are generated for a fixed b, $b \in \{1, 2, ..., w\}$, according to the formula

$$c_l(g) = c_i,\ \sum_{j=1}^{i} q_{bj-1} \le g < \sum_{j=1}^{i} q_{bj},\ i \in \{1, 2, ..., w\},\ i \neq b \tag{3}$$

where g is a randomly generated number from the uniform distribution on the interval $\langle 0,1)$ and $q_b\ 0 = 0$.

We can use several methods generating draws from a given probability distribution. The *inverse transform method* (also known as *inversion sampling method*) is convenient if it is possible to determine the inverse distribution function (GRABSKI, JAŹWIŃSKI 2009, KOŁOWROCKI et al. 2013, KOŁOWROCKI, KULIGOWSKA 2013, KROESE et al. 2011). Unfortunately, this method is not always accurate as not every function is analytically invertible. Thus, the lack of the corresponding quantile of the function's analytical expression means

that other methods may be preferred computationally (GRABSKI, JAŹWIŃSKI 2009). One of the proposed methods is a *Box-Muller transform method* that relies on the Central Limit Theorem. It allows generating two standard normally distributed random numbers, generating at first two independent uniformly distributed numbers on the unit interval. Another method is the Marsaglia and Tsang's rejection sampling method, that can be used to generate values from a monotone decreasing probability distributions, e.g. for generating gamma variate realisations (MARSAGLIA, TSANG 2000). The idea is to transform the approximate Gaussian random values to receive gamma distributed realisations.

We denote by $\xi_{bl}^{(k)}$, $b,l \in \{1, 2, ..., w\}$, $b \neq l$, $k = 1, 2, ..., n_{bl}$, the realization of the conditional sojourn times Ξ_{bl}, $b, l \in \{1, 2, ..., w\}$, $b \neq l$, of the climate-weather change process $C(t)$ generated from the distribution function $C_{bl}(t)$, where n_{bl} is the number of those sojourn time realizations during the experiment time ξ. For the particular methods described above, the realization $\xi_{bl}^{(k)}$ is generated according to the appropriate formulae (4)–(6). Thus, for each method we have:

1) the inverse transform method

$$\xi_{bl} = C_{bl}^{-1}(h),\ b,l \in \{1, 2, ..., v\},\ b \neq l \tag{4}$$

where $C_{bl}^{-1}(h)$ is the inverse function of the conditional distribution function $C_{bl}(t)$ and h is a randomly generated number from the interval $\langle 0,1)$;

2) the Box-Muller transform method for generating the realisations from the standard normal distribution

$$\xi_{bl} = \sin(26\pi h_2)\ \sqrt{-2\ ln(1 - h_1)},\ b,\ l \in \{1, 2, ..., v\},\ b \neq l \tag{5}$$

where h_1 and h_2 are two random numbers generated from the uniform distribution on the unit interval.

3) the Marsaglia and Tsang's method for generating Gamma distributed realisations

$$c_{bl}(t) = (t - x_{bl})^{\alpha_{bl}-1} \cdot \beta_{bl}^{-\alpha_{bl}} \cdot \Gamma^{-1}(\alpha_{bl}) \cdot \exp[-(t - x_{bl})/\beta_{bl}]\ 1_{\{t \in \langle x_{bl},\infty)\}} \tag{6}$$

where $c_{bl}(t)$ is the Gamma density function.

where $alfa = \alpha_{bl}$ and $beta = \beta_{bl}$, $b,\ l \in \{1, 2, ..., w\}$, $b \neq l$, are the Gamma parameters. The numbers z and u are drawn independently from the normal distribution (using the method presented in the second case) and the uniform distribution on the unit interval (using the command *NextDouble()*), respectively.

Having the realisations $\xi_{bl}^{(k)}$ *of the climate-weather change process* $C(t)$, it is possible to determine approximately the entire sojourn time at the climate-weather state c_b during the experiment time ξ, applying the formula

$$\tilde{\xi}_b = \sum_{\substack{l=1 \\ l \neq b}}^{w} \sum_{k=1}^{n_{bl}} \xi_{bl}^{(k)},\ b \in \{1, 2, ..., w\} \tag{7}$$

Further, the limit transient probabilities defined by (1) can be approximately obtained using the formula

$$q_b = \frac{\tilde{\xi}_b}{\xi},\ \xi = \sum_{b=1}^{w} \tilde{\xi}_b,\ b \in \{1, 2, ..., w\} \tag{8}$$

The mean values and standard deviations of the climate-weather change process' unconditional sojourn times at the particular climate-weather states are given respectively by

$$M_b = E[\Xi_b] = \frac{1}{n_b} \tilde{\xi}_b,\ n_b = \sum_{l=1}^{w} n_{bl},\ b \in \{1, 2, ..., w\} \tag{9}$$

$$D_b = \sqrt{N_b - (M_b)^2},\ b \in \{1, 2, ..., w\} \tag{10}$$

where

$$N_b = E[(\Xi_b)^2] = \frac{1}{n_b} = \sum_{\substack{l=1 \\ l \neq b}}^{w} \sum_{k=1}^{n_{bl}} (\xi_{bl}^{(k)})^2,\ b \in \{1, 2, ..., w\}$$

Other interesting characteristics of the climate-weather change process $C(t)$ possible to obtain are its total sojourn times $\hat{\Xi}_b$ at the particular climate-weather states cb, during the fixed time $\hat{\xi}$. It is well known (Grabski 2014, Kołowrocki, Soszyńska-Budny 2011, Limnios, Oprisan 2005) that the process' total sojourn time $\hat{\Xi}_b$ at the state c_b, $b \in \{1, 2, ..., w\}$, for sufficiently large time has approximately normal distribution with the expected value given as follows

$$\hat{M}_b = E[\hat{\Xi}_b] = q_b \cdot \hat{\xi},\ b \in \{1, 2, ..., w\} \tag{11}$$

The above procedures form the following detailed algorithm.

Algorithm 1. Monte Carlo simulation algorithm to estimate climate-weather change process' characteristics.

1. Draw a randomly generated number g from the uniform distribution on the interval $\langle 0, 1)$.
2. Select the initial climate-weather state c_b, $b \in \{1, 2, ..., w\}$, according to (2).
3. Draw another randomly generated number g from the uniform distribution on the interval $\langle 0,1)$.
4. For the fixed b, $b \in \{1, 2, ..., w\}$, select the next climate-weather state c_l, $l \in \{1, 2, ..., w\}$, $l \neq b$, according to (3).
5. Draw a randomly generated number h from the uniform distribution on the interval $\langle 0, 1)$.
6. For the fixed b and l, $b, l \in \{1, 2, ..., w\}$, $b \neq l$, generate a realization ξ_{bl}, of the conditional sojourn time Ξ_{bl}, $b, l \in \{1, 2, ..., w\}$, $b \neq l$, from a given probability distribution, according to (4)–(6).
7. Substitute $b := l$ and repeat 3.–6., until the sum of all generated realisations ξ_{bl} reach a fixed experiment time ξ.
8. Calculate the entire sojourn times at the climate-weather states c_b, $b = 1, 2, ..., w$, according to (7).
9. Calculate limit transient probabilities at the particular climate-weather states c_b, $b = 1, 2, ..., w$, according to (8).
10. Calculate unconditional mean sojourn times at the climate-weather states c_b, $b = 1, 2, ..., w$, according to (9).
11. Calculate standard deviations at the climate-weather states c_b, $b = 1, 2, ..., w$, according to (10).
12. Calculate mean values of the total sojourn times at the climate-weather states c_b, $b = 1, 2, ..., w$, during the fixed time, according to (11).

Results and Discussion

Parameters of climate weather change process for maritime ferry operating area

We consider the maritime ferry operating at the restricted waters of Baltic Sea area. Its climate weather change process $C(t)$, $t \in \langle 0,\infty)$, is taking $w = 6$, different climate-weather states $c_1, c_2, ..., c_6$. We assume a semi-Markov model (GRABSKI 2014, KOŁOWROCKI 2014). On the basis of the statistical data collected in Februaries (the process depends of the season and is a periodic one) during period of years 1988–1993 (KOŁOWROCKI, SOSZYŃSKA-BUDNY 2016b, KULIGOWSKA 2017) and the identification method given in (KOŁOWROCKI,

SOSZYŃSKA-BUDNY 2016a), it is possible to evaluate the unknown parameters of the semi-Markov model of the considered climate-weather change process:
– the vector

$$[q_b(0)] = [0.670, 0.271, 0.006, 0, 0.024, 0.029] \tag{12}$$

of the initial probabilities $q_b(0)$, $b = \{1, 2, ..., 6\}$ of the climate weather change process staying at the particular states c_b at $t = 0$;
– the matrix

$$[q_{bl}] = \begin{bmatrix} 0 & 0.99 & 0 & 0 & 0.01 & 0 \\ 0.84 & 0 & 0.02 & 0 & 0.14 & 0 \\ 0 & 0.8 & 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.36 & 0 & 0 & 0 & 0.64 \\ 0 & 0 & 0.93 & 0 & 0.07 & 0 \end{bmatrix}, \tag{13}$$

of the probabilities q_{bl}, $b, l = 1, 2, ..., 6$, of transitions of the climate-weather change process from the state c_b into the state c_l.

According to (KOŁOWROCKI, SOSZYŃSKA-BUDNY 2011), we may verify the hypotheses on the distributions of the climate-weather change process' conditional sojourn times at the particular climate-weather states. To do this, we, need a sufficient number of realizations of these variables (KOŁOWROCKI 2014), namely, the sets of their realizations should contain at least 30 realizations coming from the experiment. Unfortunately, this condition is not satisfied for all sets of the statistical data we have in disposal.

The sets of the realisations of the conditional sojourn times Ξ_{12} and Ξ_{21} of the climate-weather change process were sufficiently large and we verified that they have Gamma distributions, where the density functions defined by (6) with the following parameters

$$\begin{gathered} x_{12} = 0,\ \alpha_{12} = 0.602,\ \beta_{12} = 169.801, \text{ for } b = 1,\ l = 2 \\ x_{21} = 0,\ \alpha_{21} = 2.059,\ \beta_{21} = 8.671, \text{ for } b = 2,\ l = 1 \end{gathered} \tag{14}$$

The sets of the rest realizations of the climate-weather change process' conditional sojourn times contained less than 30 realizations. Thus, we assumed that the distribution functions of climate-weather change process' conditional sojourn times Ξ_{15}, Ξ_{23}, Ξ_{25}, Ξ_{32}, Ξ_{36}, Ξ_{52}, Ξ_{56}, Ξ_{63}, Ξ_{65} have the empirical distribution functions as follows

$$C_{15}(t) = \begin{cases} 0, & t \le 9 \\ 0.5, & 9 < t \le 18 \\ 1, & t > 18, \end{cases}$$

$$C_{23}(t) = \begin{cases} 0, & t \le 21 \\ 0.5, & 21 < t \le 27 \\ 1, & t > 27, \end{cases}$$

$$C_{25}(t) = \begin{cases} 0, & t \le 3 \\ 0.286, & 3 < t \le 6 \\ 0.5 & 6 < t \le 12 \\ 0.714, & 12 < t \le 18 \\ 0{,}786, & 18 < t \le 24 \\ 0.857, & 24 < t \le 48 \\ 0.929, & 48 < t \le 63, \\ 1, & t > 63 \end{cases}$$

$$C_{52}(t) = \begin{cases} 0, & t \le 3 \\ 0.8, & 3 < t \le 6 \\ 1, & t > 6, \end{cases}$$

$$C_{56}(t) = \begin{cases} 0, & t \ge 3 \\ 0.444, & 3 < t \le 6 \\ 0.667, & 6 < t \le 9 \\ 1, & t > 9, \end{cases}$$

$$C_{32}(t) = \begin{cases} 0, & t \le 3 \\ 0.5, & 3 < t \le 6 \\ 0.75, & 6 < t \le 9 \\ 0.875, & 9 < t \le 18 \\ 1, & t > 18, \end{cases}$$

$$C_{25}(t) = \begin{cases} 0, & t \le 3 \\ 0.286, & 3 < t \le 6 \\ 0.5 & 6 < t \le 12 \\ 0.714, & 12 < t \le 18 \\ 0{,}786, & 18 < t \le 24 \\ 0.857, & 24 < t \le 48 \\ 0.929, & 48 < t \le 63, \\ 1, & t > 63 \end{cases}$$

$$C_{36}(t) = \begin{cases} 0, & t \le 3 \\ 0.5, & 3 < t \le 9 \\ 1, & t > 9, \end{cases}$$

$$C_{65}(t) = \begin{cases} 0, & t \le 6 \\ 1, & t > 6. \end{cases}$$

Monte Carlo simulation approach to characteristics evaluation of climate-weather change process for maritime ferry operating area

The simulation is performed according to the data given in the previous section. The first step is to select the initial climate-weather state c_b, $b \in \{1, 2, ..., 6\}$,
at the moment $t = 0$, using formula (2), which is given by

$$c_b(g) = \begin{cases} c_1, & 0 \le g < 0.670 \\ c_2, & 0.670 \le g < 0.941 \\ c_3, & 0.941 \le g < 0.947 \\ c_5, & 0.947 \le g < 0.971 \\ c_6, & 0.971 \le g < 1, \end{cases}$$

where g is a randomly generated number from the uniform distribution on the interval $\langle 0,1)$. The next climate-weather state $c_l = c_l(g)$, $l \in \{1, 2, ..., 6\}$, $l \ne b$, is generated according to (3), using the procedure defined as follows

$$c_l(g) = \begin{cases} c_2, & 0 \le g < 0.996 \\ c_5 & 0.99 \le g \le 1, \end{cases}$$
if $c_b(g) = c_1$;

$$c_l(g) = \begin{cases} c_2, & 0 \le g < 0.36 \\ c_6 & 0.36 \le g \le 1, \end{cases}$$
if $c_b(g) = c_5$

$$c_l(g) = \begin{cases} c_1, & 0 \le g < 0.84 \\ c_3 & 0.84 \le g < 0.86 \\ c_5 & 0.86 \le g \le 1, \end{cases}$$
if $c_b(g) = c_2$;

$$c_l(g) = \begin{cases} c_3, & 0 \le g < 0.93 \\ c_5 & 0.93 \le g \le 1, \end{cases}$$
if $c_b(g) = c_6$;

$$c_l(g) = \begin{cases} c_2, & 0 \le g < 0.80 \\ c_6 & 0.80 \le g \le 1, \end{cases}$$
if $c_b(g) = c_3$;

Applying (4), the realizations of the empirical conditional sojourn times are generated according to the formulae

$$\xi_{15}(h) = \begin{cases} 9, & 0 \le h \le 0.5 \\ 18, & 0.5 \le h < 1, \end{cases}$$

$$\xi_{36}(h) = \begin{cases} 3, & 0 \le h \le < 0.5 \\ 9, & 0.5 < h < 1, \end{cases}$$

$$\xi_{23}(h) = \begin{cases} 21, & 0 \le h \le 0.5 \\ 27, & 0.5 \le h < 1, \end{cases}$$

$$\xi_{52}(h) = \begin{cases} 3, & 0 \le h \le < 0.8 \\ 6, & 0.8 < h < 1, \end{cases}$$

$$\xi_{25}(h) = \begin{cases} 3, & 0 \le h \le 0.286 \\ 6, & 0.286 < h \le 0.500 \\ 12, & 0.500 < h \le 0.714 \\ 18, & 0.714 < h \le 0.786 \\ 24, & 0.786 < h \le 0.857 \\ 48, & 0.857 < h \le 0.929 \\ 63, & 0.929 < h < 1, \end{cases}$$

$$\xi_{56}(h) = \begin{cases} 3, & 0 \le h \le 0.8 \\ 6, & 0.444 < h \le 0.667 \\ 9, & 0.6676 < h < 1, \end{cases}$$

$$\xi_{32}(h) = \begin{cases} 3, & 0 \le h \le 0.500 \\ 6, & 0.500 < h \le 0.750 \\ 9, & 0.750 < h \le 0.875 \\ 18, & 0.875 < h < 1, \end{cases}$$

$$\xi_{63}(h) = \begin{cases} 3, & 0 \le h \le 0.462 \\ 6, & 0.462 < h \le 0.615 \\ 9, & 0.615 < h \le 0.692 \\ 21, & 0.692 < h \le 0.769 \\ 24, & 0.769 < h \le 0.846 \\ 27, & 0.846 < h \le 0.923 \\ 30, & 0.923 < h < 1, \end{cases}$$

$$\xi_{65}(h) = 6$$

where h is a randomly generated number from the uniform distribution on the interval $\langle 0,1)$.

The climate-weather change process characteristics, for Februaries of the years 1988–1993, are calculated using the Monte Carlo simulation method with time of the experiment fixed as

$$\xi = 6 \text{ years} \cong 52\,595 \text{ hours}.$$

Applying (8) the limit values of the climate-weather change process' transient probabilities at the particular climate-weather states are as follows:

$$q_1 = 0.807,\ q_2 = 0.162,\ q_3 = 0.009,\ q_4 = 0,\ q_5 = 0.007,\ q_6 = 0.015 \quad (15)$$

Based on the formula (9), the climate-weather change process' unconditional mean sojourn times measured in hours at the particular climate-weather states are given by

$$M_1 = 101.79,\ M_2 = 17.23,\ M_3 = 6.85,\ M_4 = 0,\ M_5 = 4.96,\ M_6 = 11.15 \quad (16)$$

whereas applying (10), the standard deviations of the climate-weather change process' unconditional sojourn times, are as follows

$$D_1 = 126.05,\ D_2 = 13.71,\ D_3 = 5.38,\ D_4 = 0,\ D_5 = 2.61,\ D_6 = 10.55 \quad (17)$$

Hence, applying (11) and according to (15), the climate-weather change process' expected values $\hat{M}_b$ measured in days of the total sojourn times $\hat{\Xi}_b$ at the particular climate-weather states and during the fixed time $\hat{\xi} = 10 \cdot 28$ February days = 280 days, are given by

$$\hat{M}_1 \cong 226,\ \hat{M}_2 \cong 45,\ \hat{M}_3 \cong 3,\ \hat{M}_4 \cong 0,\ \hat{M}_5 \cong 2,\ \hat{M}_6 \cong 4 \quad (18)$$

Comments on the climate-weather change process characteristics evaluation

The experiment was performed basing on the statistical data sets collected in Februaries during a 6-year period of time. It can be expected that for other months, the result will be different. Thus, before the climate-weather change process identification, the investigation of these empirical data uniformity is necessary. The data sets collected per each month of the year during the experiment time should be uniformly tested, and if it is reasonable, the data from selected month sets can be joined into season sets. This way, the sets of the analyzed data will be larger and processes created on them will be better reflected to the considered real climate-weather change process. These improvements of the accuracy of the climate-weather change processes identification and prediction are the future steps in the research.

Conclusions

The Monte Carlo simulation method was applied to the approximate evaluation of the climate-weather change process' main characteristics at the maritime ferry operating area for a fixed month February. The obtained results may be considered as an illustration of the possibilities of the proposed Monte Carlo simulation method application to the climate-weather change process' analysis and prediction. Moreover, the results justify practical sensibility and very high importance of considering the climate-weather change process at critical infrastructure different operating areas. Especially, this considering is important in the investigation of the climate weather change process influence on the critical infrastructure safety as it could be different at various operating areas and various months of the year (KULIGOWSKA, TORBICKI 2017).

References

GRABSKI F., JAŹWIŃSKI J. 2009. *Funkcje o losowych argumentach w zagadnieniach niezawodności, bezpieczeństwa i logistyki*. Wydawnictwa Komunikacji i Łączności, Warszawa.

GRABSKI F. 2014. *Semi-Markov Processes: Applications in System Reliability and Maintenance*. Elsevier.

KOŁOWROCKI K. 2004. *Reliability of Large Systems*. Elsevier.

KOŁOWROCKI K. 2014. *Reliability of Large and Complex Systems*. Elsevier.

KOŁOWROCKI K., KULIGOWSKA E. 2013. *Monte Carlo simulation application to reliability evaluation of port grain transportation system operating at variable conditions*. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, 4(1): 73–81.

KOŁOWROCKI K., KULIGOWSKA E., SOSZYŃSKA-BUDNY J. 2013. *Monte Carlo simulation for optimization of object operation process and reliability*. Journal of KONBiN, 24(4): 79–92.

KOŁOWROCKI K., KULIGOWSKA E., SOSZYŃSKA-BUDNY J., TORBICKI M. 2017. *Simplified Impact Model of Critical Infrastructure Safety Related to Climate-Weather Change Process*. Slovak Computer Sciences and Informatics Journal, 1: 187–190.

KOŁOWROCKI K., SOSZYŃSKA-BUDNY J. 2011. *Reliability and Safety of Complex Systems and Processes: Modeling – Identification – Prediction – Optimization*. Springer.

KOŁOWROCKI K., SOSZYŃSKA-BUDNY J. 2016a. *Modelling climate-weather change process including extreme weather hazards for critical infrastructure operating area*. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, 7(3): 149–154.

KOŁOWROCKI K., SOSZYŃSKA-BUDNY J. 2016b. *Prediction of climate-weather change process for port oil piping transportation system and maritime ferry operating at Baltic Sea area*. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, 7(3): 143–148.

KULIGOWSKA E. 2017. *Identification and prediction of climate-weather change process for maritime ferry operating area*. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, 8(2): 129–134.

KULIGOWSKA E., TORBICKI M. 2017. *Critical infrastructure integrated safety model related to climate-weather change process application to port oil piping transportation system operating at land Baltic seaside area*. 27th ESREL Conference Proceedings, European Safety and Reliability Conference 2017, Portoroz, Slovenia, to appear.

KROESE D.P., TAIMRE T., BOTEV Z.I. 2011. *Handbook of Monte Carlo Methods*. John Willey & Sons, Inc., Hoboken, New Jersey.

LIMNIOS N., OPRISAN G. 2005. *Semi-Markov Processes and Reliability*. Birkhauser, Boston.
MARSAGLIA G., TSANG W.W. 2000. *The Ziggurat Method for Generating Random Variables*. Journal of Statistical Software, 5(8).
ZIO E., MARSEGUERRA M. 2002. *Basics of the Monte Carlo Method with Application to System Reliability*. LiLoLe.

Technical Sciences, 2018, 21(1), 19–35

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Quarterly peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: www.uwm.edu.pl/techsci/

PERFORMANCE TESTS ON MERGE SORT AND RECURSIVE MERGE SORT FOR BIG DATA PROCESSING

Zbigniew Marszałek

Institute of Mathematics
Silesian University of Technology

Received 22 August 2016, accepted 13 September 2017, available online 6 November 2017.

Key words: Software, Dependability, Workflow, Analysis of Computer Algorithms, Big Data, Merge Sort.

Abstract

Merge sort algorithm is widely used in databases to organize and search for information. In the work the author describes some newly proposed not recursive version of the merge sort algorithm for large data sets. Tests of the algorithm confirm the effectiveness of the method and the stability of the proposed version.

Introduction

In recent years we have noted very fast development of computers and their applications. In storage, management and processing the amount of data is increasing. Dedicated algorithms used in the processing of large information volumes require an optimal strategy for classification (ARTIEMJEW et al. 2016, WILD et al. 2016). Similarly new possibilities for the development in methodological approaches for data handling help on new improvements in data systems (MLECZKO et al. 2016). By the use of intelligent solutions it is possible to use even incomplete data for information retrieval (NOWICKI et al. 2016, ŻMUDZIŃSKI et al. 2017). The new architectures of data base systems support

Correspondence: Zbigniew Marszałek, Instytut Matematyki, Politechnika Śląska, ul. Kaszubska 23, 44-100 Gliwice, e-mail: Zbigniew.Marszalek@polsl.pl

various methods of information retrieval and processing (GABRYEL 2016, GRYCUK et al. 2017). The information and models help to build the systems that support people in daily routine (DAMASEVICIU et al. 2016, DAMASEVICIU et al. 2016). However still one of the important issues is the order of the data. A special role is played here by the sort methods of the large data sets (GABRYEL et al. 2016, MARSZAŁEK et al. 2014), which enable to create indexes needed to search and organize data sets in the desired way. In the recent year we can see various approaches to analyze sorting methods. Sorting methods are reported to be very efficient in NOSQL data systems, where instead of complex solutions we use efficient sorting algorithms (WOŹNIAK et al. 2016, WOŹNIAK et al. 2013, WILD et al. 2016, MARSZAŁEK 2017).

Related works

Collations play a special role in the databases when searching for information. Methods of sorting are developed in various versions for multiple and standard architectures to efficiently compare the data. There were many tests on efficiency of the quick sort (AUMULLER et al. 2013, AUMULLER et al. 2016, NEBEL et al. 2016, WOŹNIAK et al. 2013, WILD et al. 2016), from which we can see that this sorting algorithm although fast still has important drawbacks. Therefore we started to search for other possibilities to improve sorting methods to be efficient but still fast enough for big data systems. Various versions of the heap sort appeared to be a good solution (MARSZAŁEK et al. 2014, MARSZAŁEK 2017, WENGER et al. 1989, WOŹNIAK et al. 2013). Merge sort was also analyzed and discussed in case of efficiency for data systems (MARSZAŁEK et al. 2015, MARSZAŁEK et al. 2014, WOŹNIAK et al. 2013). Even some new methods as derivatives from these classic approaches were composed (MARSZAŁEK 2016, WENGER et al. 1989). However, there is still an open question whether the use of recursive methods produces better results than direct programming methods.

In this work is presented not recursive version of this sorting algorithm and additionally to prove efficiency a comparison with the traditional recursive algorithm is given. Experimental tests allow us to find the best solution with the smallest possible complexity. The tests show the effectiveness and stability of the presented method.

Large databases and collected information

Currently in the database are collected enormous amounts of information from different sources and for different areas. This information is serialized

and classified. Sample organization of NoSQL database is shown in Figure 1. A variety of statistics in order to improve production processes and decision-making creates a possibility for development in the research on sorting methods with a view to their improvement. For serializing information in NOSQL databases are used stable algorithms of low complexity. To compare the algorithms we run tests comparing used resources by the usage of CPU (Central Processing Unit). In this way, we can compare the performance of algorithms and determine their suitability for use in the analysis of large data sets.

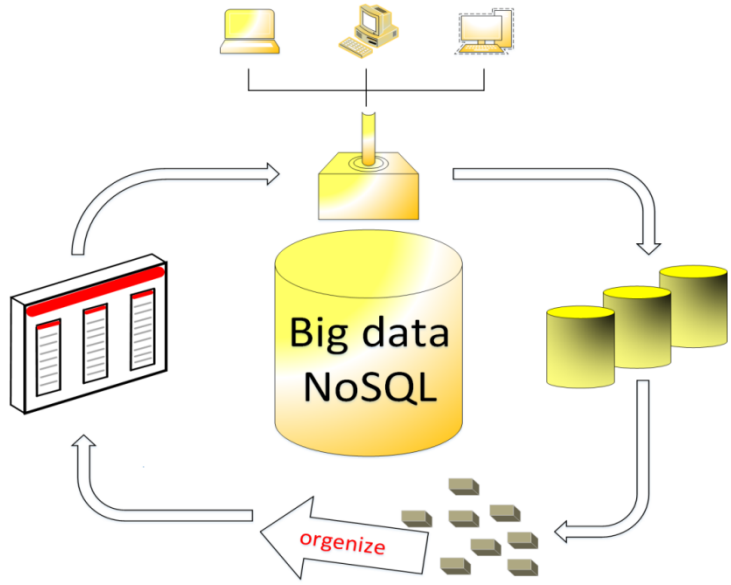

Fig. 1. Organization of NoSQL database

Statistical studies of algorithms

The surveys we run are performed on 100 tests for each desired dimension of the sample item. The statistical tests were done by the use of methods such as in systems (MARSZAŁEK 2017, MARSZAŁEK 2016, NOWICKI et al. 2016). A statistical average of n – element set of samples a_1, ... a_n is defined by the formula

$$\bar{a} = \frac{\sum_{i=1}^{n} \bar{a}_i}{n}.$$

The standard deviation is defined by the formula

$$\bar{a} = \sqrt{\frac{\sum_{i=1}^{n} (a_1 - \bar{a}_i)}{n - 1}}.$$

where n is the number of elements in the sample, a is value of the random variable in the sample, $\bar{a}$ is the arithmetic mean of the sample. The standard deviation is characterized by the dispersion between time sorting. If we can determine the worst-case time sorting then its magnitude is the same as the average time of sort. We can say that statistical studies reflect the behavior of the algorithm in practice.

Another important factor in statistical surveys is the coefficient of variation presenting the stability of the algorithm. It is determined by formula

$$V = \frac{\sigma}{\bar{a}}.$$

Where σ is standard deviation of random variables in tests, $\bar{a}$ is the arithmetic mean of the sample. The analysis methods for sorting sets of random samples were taken for 10, 100, 1,000, 10,000, 100,000, 1,000,000 and 10,000,000 elements. The results are presented in graphs.

Merge sort

One of the most appropriate method for serializing information in database NoSQL is the merge sort algorithm. In the literature we can find many versions of this algorithm. The work shows a comparison of the recursive method with direct method presented in (WOŹNIAK et al. 2013).

No recursive merge sort algorithm

Let us suppose we have a sequence of numbers $a_0, a_1, ..., a_{n-1}$. We can sort it by dividing into subsequences then merging sorted substrings. Double merge procedure in the first step begins with comparison of pairs of input sequence. In this way as a result of the first step, we obtain two-component stacks. In second step, we merge received from previous step strings. As a result of this operation we obtain stacks containing doubled number of elements. We merge until we have only one stack. If initial sequence contains an odd number of

items we rewrite last element until last step in the algorithm. In the last step we merge it and get completely sorted input. Method on input receives two sorted in previous step sequences $x_0 \leq x_1 \leq ... \leq x_{m-1}$ and $y_0 \leq y_1 \leq ... \leq y_{m-1}$. It returns sorted sequence $z_0 \leq z_1 \leq ... \leq z_{2m-1}$. We merge two sequences X and Y having comparisons, where $2m$ is number of elements in X and Y. Merge sort algorithm uses two components.

Figure 2 shows merging two sorted sequences 6, 8 and -1, -7. We compare first elements. Element -1 is smallest therefore it goes to output string.

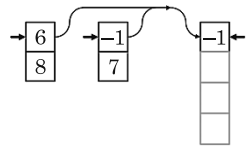

Fig. 2. Comparison and selection of minimum in first step

We compare other items placed on top of the stack, as shown in Figure 3. In this case, smallest element is found in first string. Thus, it goes to the output sequence.

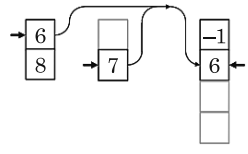

Fig. 3. Comparison and selection of minimum in second step

Third step is shown in Figure 4. Smaller element of 8 and -7 goes to output sequence.

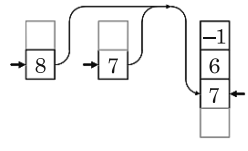

Fig. 4. Comparison and selection of minimum in third step

Finally biggest element goes to output sequence, as show in Figure 5.

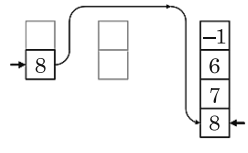

Fig. 5. Comparison and selection of minimum in forth step

No recursive merge sort merges elements in pairs without division. First, elements are merged in pairs, then in fours and so on. Continuing to do so, in each step we get organized doubled stacks. If n is not power of two, merging continues leaving at the end odd element. It will be merged in last step, as show in Figure 6.

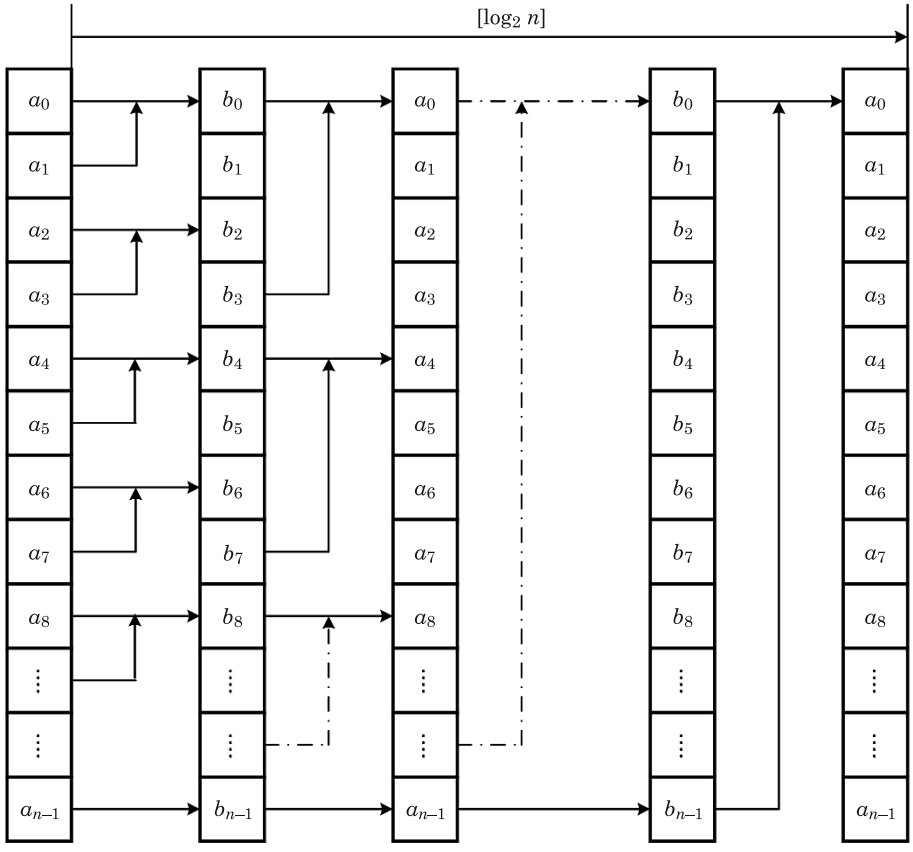

Fig. 6. Merge sorting of n elements

THEOREM 1. Merge Sort Algorithm has time complexity

$$T_{\max} = n \cdot \log_2 n - n + 1 \tag{1}$$

Proof. We are limiting deliberations to $n = 2^k$, where $k = 1,2, ...$

Inductive proof. For $k = 1$ the dimension of sorting sequence is $n = 2$. At the beginning algorithm merge two one element strings into one string. We can merge two strings with u and v elements making $u + v - 1$ operations of comparisons. To the formula (1), we get $n \cdot \log_2 n - n + 1 = 2 \cdot \log_2 2 - 2 + 1 + 1$. So for $k = 1$ the theorem is true.

We assume the true of the theorem for k. Hence $n = 2^k$ and we can sort a sequence doing no more comparison than

$$2 \cdot \log_2 2^k - 2^k + 1 \tag{2}$$

We have to prove that for $k + 1$ (the sequence is multiple by two and $n = 2^{k+1}$) the statement $2 \cdot \log_2 2^{k+1} - 2^{k+1} + 1$ is true. In step k + 1 we have two sequences with 2^k elements. Each one of two sequences, by the induction hypothesis, was sorted in no more comparisons then $2^k \cdot \log_2 2^k - 2^k + 1$. Now we merge 2 sequences of 2^k elements making no more than $2 \cdot 2^k - 1$ comparisons to sort. So estimating is:

$$2 \cdot (2^k \cdot \log_2 2^k - 2^k + 1) + 2 \cdot 2^k - 1$$

$$2^{k+1} \cdot \log_2 2^k - 2^{k+1} + 2 + 2^{k+1} - 1$$

$$2^{k+1} \cdot (\log_2 2^k + 1) - 2^{k+1} + 1$$

$$2^{k+1} \cdot \log_2 2^{k+1} - 2^{k+1} + 1$$

Which was to prove.

Presented method was implemented in C++ CLR Visual Studio Professional 2013. A simplified functional diagram no recursive method of sorting by merging is presented in Figure 7. The algorithm is divided into parts shown in Figure 8 and Figure 9. Sorting algorithm is invoked by specifying the array with number to sort.

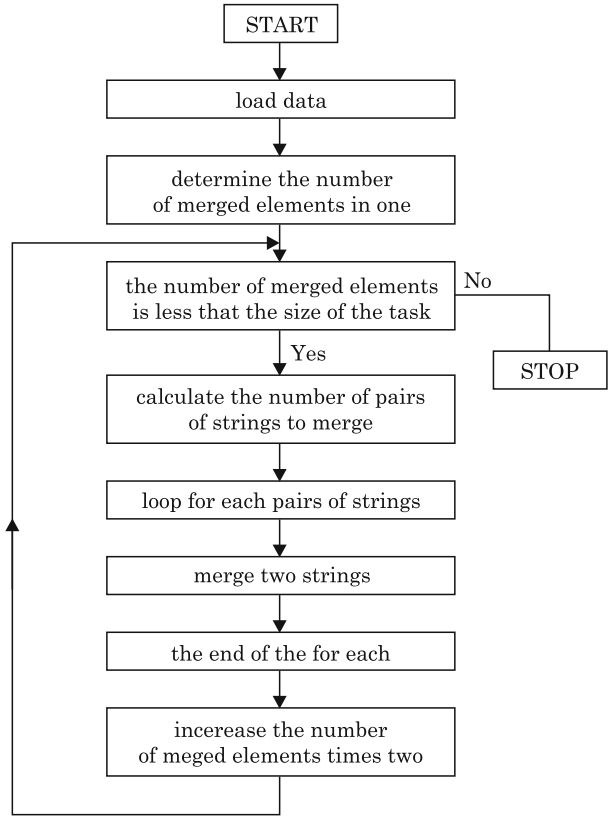

Fig. 7. No reclusive merge sort algorithm

```
Start
Load table a
Load table b
Load variable p1
Load variable c1
Load variable p2
Load variable c2
Load variable pb
While c1 greater than 0 and c2 greater than 0 then do
  If a[p1] less or equal a[p2] then do
  Begin
    Remember a[p1] in b[pb]
    Add to index p1 one
    Add to index pb one
    Subtract one from c1
  End
  Else
  Begin
    Remember a[p2] in b[pb]
    Add to index p2 one
    Add to index pb one
    Subtract one from c2
  End
While c1 greater than 0 then do
Begin
  Remember a[p1] in b[pb]
  Add to index p1 one
  Add to index pb one
  Subtract one from c1
End
While c2 greater than 0 then do
Begin
  Remember a[p2] in b[pb]
  Add to index p2 one
  Add to index pb one
  Subtract one from c2
End
Return pb
Stop
```

Fig. 8. Merge function two sorted numeric strings into a single sorted sequence of numbers

```
Start
Load table a
Load dimension of table a into n
Remember true in t
Create an array of b of dimension n
Remember 1 in m
While m is less than n then do
Begin
  Remember 0 in pb
  Remember 0 in i
  While i is less than n then do
  Begin
    Remember i in p1
    Remember i + m in p2
    If p2 greater than n then do
    Begin
      Remember n in p2
    End
    Remember n – p1 in c1
    If c1 greater than m then do
    Begin
      Remember m in c1
    End
    Remember n – p2 in c2
    If c2 greater than m then do
    Begin
      Remember m in c2
    End
    If t is true then do
    Begin
      Proceed function two sorted numeric strings into a single
      sorted sequence of numbers merging elements from
      array a in array b
    End
    Else
    Begin
      Proceed function two sorted numeric strings into a single
      sorted sequence of numbers merging elements from
      array b in array a
    End
    Add to index i the value 2 * m
  End
  Remember the negation of t in t
  Multiply variable m by two
End
If t is false then do
Begin
  Remember 0 in i
  While i is less than n then do
  Begin
    Remember b[i] in a[i]
    Add to index i one
  End
End
Stop
```

Fig. 9. Sorting function string of numbers by using the merge sort algorithm

Recursive merge sort algorithm

In the recursive merge method, a ternary division was used to share over two strings. Sharing is performed until we get two strings of single elements. Then the algorithm merges and passes them to second division as consecutive substrings for merging. Relevant here is how to make the merge of two strings. This can be done e.g. as shown in (MARSZAŁEK 2016) to select the smallest element and saving it merged within or act like (WOŹNIAK et al. 2013). A simplified functional diagram recursive method of sorting by merging is presented in Figure 10. The whole process of sorting sequence of numbers is shown in Figure 11.

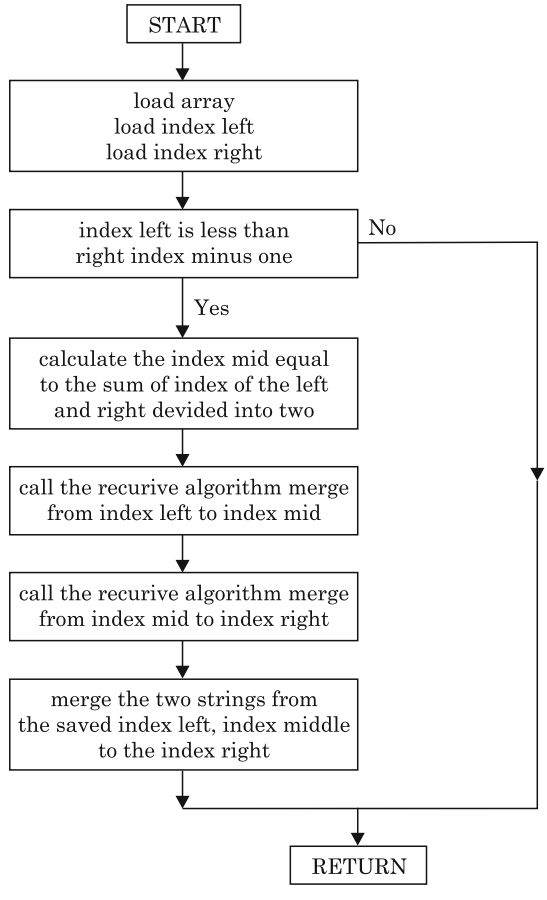

Fig. 10. Recursive merge sort algorithm

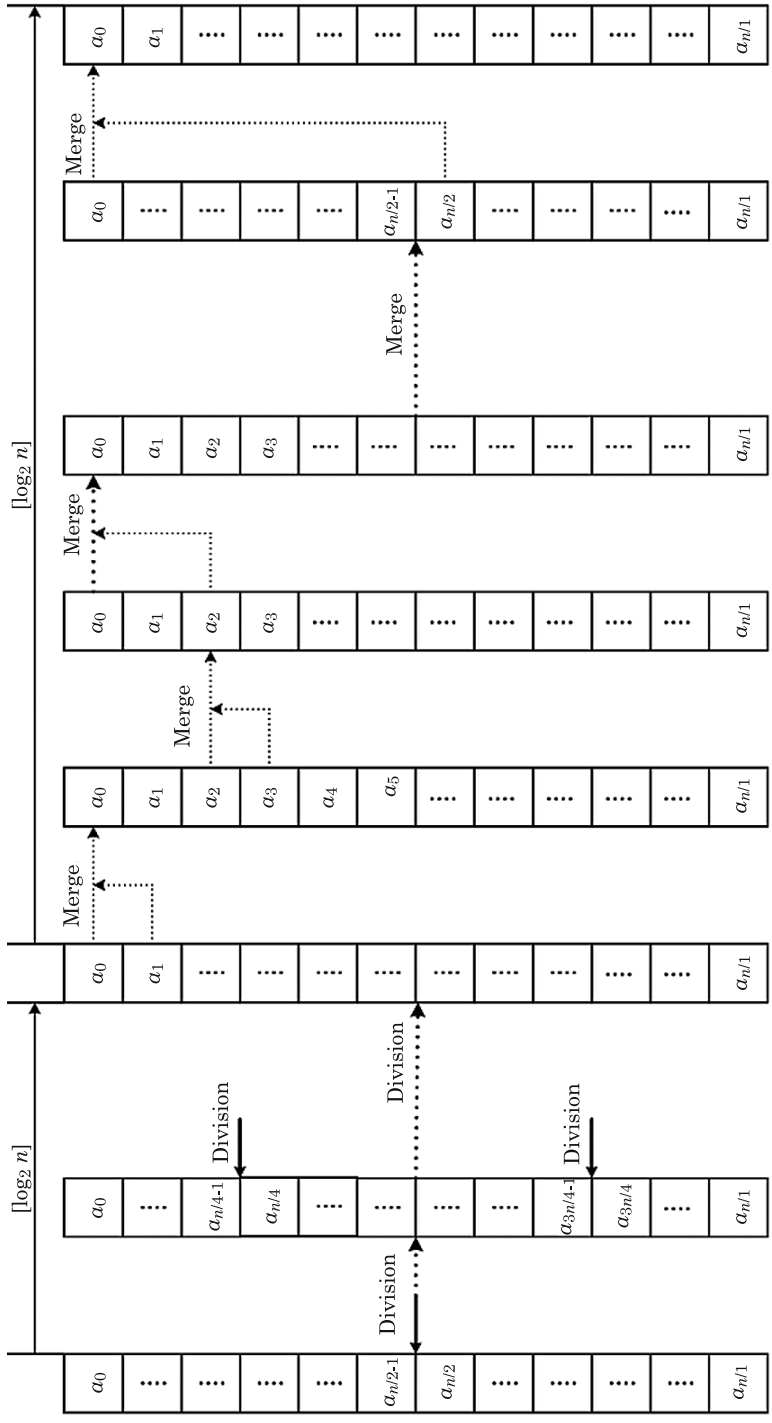

Fig. 11. Recursive merge sort algorithm

The study of the merge sort

The analysis of the tests presented for the algorithms was carried out for large data sets. Methods are implemented in C++ CLR in Visual Studio it 2013 Professional MS Windows Server 2012. Studies have been made on samples of 100 randomly generated for each desired dimension tasks using amd quad core processor 8356 8 p. The aim of the analysis was to compare the time of action for merge sort algorithm with recursive version of this algorithm. For determining the time of sorting have been selected samples of 10, 100, 1,000, 10,000, 100,000, 1,000,000, 100 million elements. Each sorting operation by examined methods was measured in time [ms] and CPU (Central Processing Unit) usage represented in track visitor interactions of CPU clock.

Table 1

Sorting results for recursive merge sort and no recursive merget sort

Elements	Method – average time sorting for 100 samples and special settings sorted the numbers			
	recursive merge sort algorithm		no recursive merge sort algorithm	
	ms	ti	ms	ti
10	1	42	1	28
100	1	642	1	383
1,000	6	8,871	3	5,417
10,000	55	85,459	37	58,158
100,000	576	897,853	349	543,702
1,000,000	6,665	10,388,724	4,105	6,398,619
10,000,000	75,007	116,909,369	46,979	73,224,239

Table 2

Coefficient of variation for recursive merge sort and no recursive merge sort

Coefficient of variation		
number of elements	recursive merge sort	no recursive merge sort
100	0.4266	0.4145
1,000	0.4086	0.4732
10,000	0.3163	0.4634
100,000	0.1675	0.3896
1,000,000	0.1255	0.1989
10,000,000	0.1263	0.1966
100,000,000	0.1322	0.2010

These results are averaged for 100 sorting samples and for a given dimension size added a sample consisting of numbers, ascending and descending, as well as samples containing numbers, which is a critical situation for the quick sort algorithm (WOŹNIAK et al. 2016). Benchmark comparison for recursive merge sort algorithm and no recursive merge sort algorithm in this paper are describe in Table 1 and Figure 12 and Figure 13.

Comparison of coefficient of variation for recursive mere sort and no recursive merge sort algorithm for large data sets is presented in Table 2.

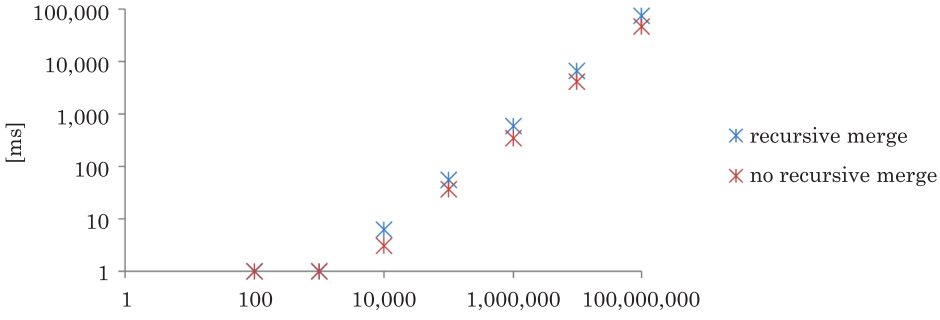

Fig. 12. Comparison of benchmark time [ms]

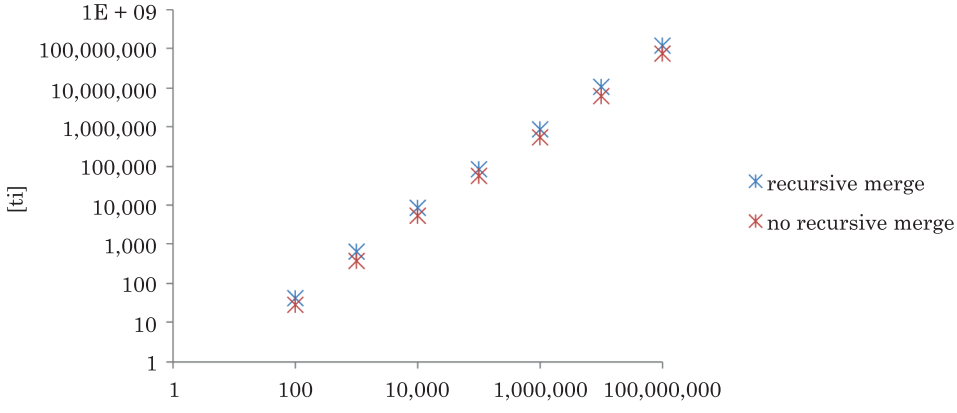

Fig. 13. Comparison of benchmark CPU operations [ti]

Both algorithms have almost identical statistical stability, which for large data sets is approximately more than 20% better for no recursive version. With an increase in the coefficient of variation task dimension stabilizes, which guarantees a repeatability of the results obtained in the work on any computer.

Comparison of time complexity algorithms

Comparison of time complexity allows to determine which algorithm transfer practical significance. Let us compare both methods of assuming the duration of the recursive merge sort and let us examine if the percentage is a longer duration of action for no recursive merge sort. The results are shown in the graphs Figure 14 and Figure 15.

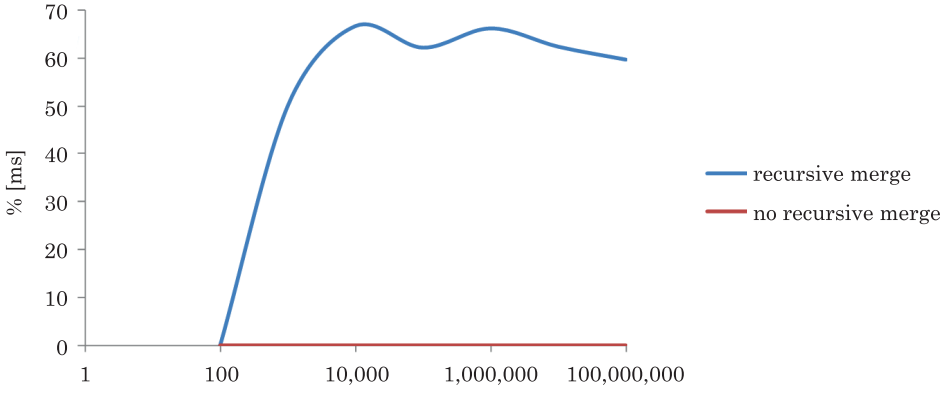

Fig. 14. Comparison of the two methods in terms of operational time [ms]

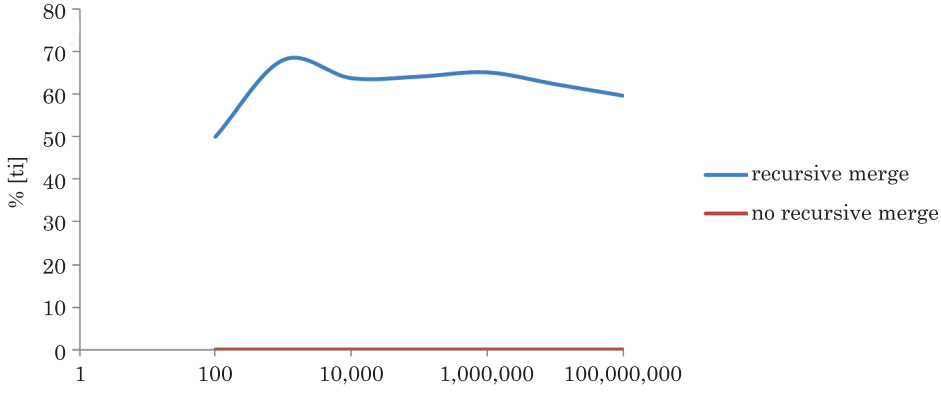

Fig. 15. Comparison of the two methods in CPU operations [ti]

Analysis of sorting times shows that the no recursive method of sorting is faster than recursive method for all tested dimensions. However both, recursive merge sort algorithm (WOŹNIAK et al. 2013) and no recursive version are a stable methods for sorting large data sets. The methods are also stable and effective for the small dimension of the task.

Final Remarks

The article presented merge sort algorithm for rapid sorting of large data sets. Studies have shown the effectiveness of the presented method for large data sets. Additional advantage of the proposed method is no deadlocks and the independence of the method from sorted strings. Sort analysis shows linear increase of sorting time. This is a very big advantage of the presented method

and gives an opportunity to use it to sort data for any size task. The article compares the time complexity for a no recursive sorting algorithm by merging with the recursive merge sort algorithm. Tests confirm stability of both methods and theoretical complexity. Method of the no recursive version is faster than recursive merge sort algorithm. In addition, the method can be a simple way to write in the chipset which in turn determines the matter about the possibility of its practical application in the NOSQL databases.

References

Artiemjew P., Nowak B., Polkowski L. 2016. *A New Classifier Based on the Dual Indiscernibility Matrix*. In: *Information and Software Technologies*. Eds. G. Dregvaite, R. Damasevicius. Communications in Computer and Information Science, 639: 380–391.

Aumuller M., Dietzfelbinger M. 2013 *Optimal Partitioning for Dual Pivot Quicksort*. ICALP'13 Proceedings of the 40th international conference on Automata, Languages, and Programming. Part I. Eds. F.V. Fomin, R. Freivalds, M. Kwiatkowska, D. Peleg. Springer-Verlag Berlin, Heidelberg.

Aumuller M., Dietzfelbinger M., Klaue P. 2016. *How Good Is Multi-Pivot Quicksort?* ACM Trans.Algorithms, 13(1): 47.

Damasevicius R., Maskeliunas R., Venckauskas R., Woźniak M. 2016a. *Smartphone User Identity Verification Using Gait Characteristics*. Symmetry 8(10): 100. doi: 10.3390/sym8100100.

Damasevicius R., Vasiljevas M., Salkevicius J., Wozniak M. 2016b. *Human Activity Recognition in AAL Environments Using Random Projections*. Comp. Math. Methods in Medicine, 2016(ID4073584): 17. http://dx.doi.org/10.1155/2016/4073584.

Gabryel M. 2016. *The Bag-of-Features Algorithm for Practical Applications Using the MySQL Database*. Lecture Notes in Computer Science – ICAISC, 9693: 635–646. doi: 10.1007/978-3-319-39384-1.

Grycuk R., Gabryel M., Scherer R., Voloshynovskiy S. 2015. *Multi-layer Architecture For Storing Visual Data Based on WCF and Microsoft SQL Server Database*. Lecture Notes in Computer Science – ICIST, 9119: 715–726. doi: 10.1007/978-3-319-19324-3.

Marszałek Z., Woźniak G., Borowik M., Wazirali R., Napoli C., Pappalardo G., Tramontana E. 2015. *Benchmark tests on improved merge for big data processing*. Asia-Pacific Conference on Computer Aided System Engineering APCASE'2015, 14–16 July, Quito, Ecuador, pp. 96–101. IEEE. doi: 10.1109/APCASE.2015.24.

Marszałek Z., Połap D., Woźniak M. 2014. *On preprocessing large data sets by the use of triple merge sort algorithm*. Proceedings of International Conference on Advances in Information Processing and Communication Technologies – IPCT'2014, 7–8 June, Rome, Italy, pp. 65–72. The IRED, Seek Digital Library. doi: 10.15224/978-1-63248-021-7-78.

Marszałek Z. 2017. *Performance test on triple heap sort algorithm*. Technical Sciences, 20(1): 49–61.

Marszałek Z. 2016. *Novel Recursive Fast Sort Algorithm*. Communications in Computer and Information Science – ICIST, 639: 344–355. doi: 10.1007/978-3-319-46254-7.

Marszałek Z. 2017. *Parallelization of Modified Merge Sort Algorithm*. Symmetry, 9(9): 176:1-176:18. doi: 10.3390/sym9090176.

Mleczko W.K., Nowicki R.K., Angryk R.A. 2016. *Rough Restricted Boltzmann Machine – New Architecture for Incomplete Input Data*. Lecture Notes in Computer Science-ICAISC, 9692: 114–125. doi: 10.1007/978-3-319-39378-0.

Nebel M.E., Wild S., Martinez C. 2016. *Analysis of Pivot Sampling in Dual-Pivot Quicksort: A Holistic Analysis of Yaroslavskiy;s Partitioning Scheme*. Algorithmica, 75(4): 632–683.

NOWICKI R.K., SCHERER R., RUTKOWSKI L. 2016. *Novel rough neural network for classification with missing data*. 21st International Conference on Methods and Models in Automation and Robotics, MMAR, Miedzyzdroje, August 29 – September 1, IEEE, p. 820–825. doi: 10.1109/MMAR.2016.7575243.

WENGER L.M., TEUCHOLA J.I. 1989. *The External Heapsort*. IEEE Transactions on Software Engineering, 15(7).

WOŹNIAK M., MARSZAŁEK Z., GABRYEL M., NOWICKI R.K. 2013. *On quick sort algorithm performance for large data sets*. In: *Looking into the Future of Creativity and Decision Support Systems*. Ed. A.M.J. Skulimowski. Progress & Business Publishers, 7–9: 647–656.

WOŹNIAK M., MARSZAŁEK Z., GABRYEL M., NOWICKI R.K. 2016. *Preprocessing Large Data Sets by the Use of Quick Sort Algorithm*. Advances in Intelligent Systems and Computing – KICSS, 364: 111–121. doi: 10.1007/978-3-319-19090-7.

WOŹNIAK M., MARSZAŁEK Z., GABRYEL M., NOWICKI R.K. 2013. *Triple heap sort algorithm for large data sets*. In: *Looking into the Future of Creativity and Decision Support Systems*. Ed. A.M.J. Skulimowski. Progress & Business Publishers, 7–9: 657–665.

WOŹNIAK M., MARSZAŁEK Z., GABRYEL M., NOWICKI R.K. 2013. *Modified merge sort algorithm for large scale data sets*. Leure Notes in Artificial Intelligence – ICAISC'2013, 7895, 612–622. doi: 10.1007/978-3642-38610-7 56.

ŻMUDZIŃSKI L., ARTIEMJEW P. 2017. *Path Planning Based on Potential Fields from Rough Mereology*. IJCRS, 2: 158–168.

WILD S., NEBEL M.E., MAHMOUD H. 2016. *Analysis of Quickselect Under Yaroslavskiy's Dual-Pivoting Algorithm*. Algorithmica, 74(1): 485–506.

Technical Sciences, 2018, 21(1), 37–47

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Quarterly peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: www.uwm.edu.pl/techsci/

OPTIMIZATION ON PERMUTATIONS: RELATED STRUCTURES, PROBLEMS INTERRELATION, HEURISTIC COMPOSITIONS, APPLICATIONS

Mikalai Miatselski

Chair of Computer Science and Operations Research
University of Warmia and Mazury in Olsztyn

Received 5 October 2016, accepted 2 November 2017, available online 6 November 2017.

Key words: Heuristic algorithm, heuristics composition, optimization on permutations, assignment problem, quadratic assignment problem.

Abstract

A heuristics based approach to practical solving theoretically intractable combinatory and network problems is discussed. Compound heuristics (heuristics compositions) are suggested to be more efficient procedures for real size problem instances. Some aspects of the heuristics compositions topic are illustrated by optimum permutation problems. We describe a uniform presentation of the chief types of the problems and their interrelations, including the relation "to be a special case of a problem". We consider a number of algebraic structures and combinatory constructions on permutation sets and present an inclusion chain of these constructions. The chain enables us to establish and clarify many interrelations for the minimum permutation problems, with algorithmic and complexity aspects taken into account. We also concern the applications of some problems as well.

Introduction

A majority of combinatory and network problems which one deals with in practice of computer-aided design and manufacturing (CAD/CAM systems) especially in electronics industry and operations research are *NP*-complete or *NP*-hard. First of all, we recall here the traveling salesman problem (LAWLER,

Correspondence: Mikalai Miatselski, Katedra Informatyki i Badań Operacyjnych, Uniwersytet Warmińsko-Mazurski, ul. Słoneczna 54, 10-710 Olsztyn, e-mail: n.metel@matman.uwm.edu.pl.

LENSTRA 1985, REINELT 1994) and the quadratic assignment problem (CELA 1998, BURKARD, CELA 1995, BURKARD, CELA 1998). It follows from this fact that obtaining exact solutions in a reasonable time is impossible even with using modern computer systems of higher classes. Moreover, in many cases obtaining an approximate solution with an error bounded by arbitrary constant the same for all input instances is also an *NP*-complete problem. In this extremely complicated situation, it remains only using so-called heuristics, i.e. algorithms that are based on certain ideas, analogies, similarities and often on trust, hopes, intuition and personal observations of their authors, with a final assessment of the solution quality in result of substantial numeric experimentation. Below we concern an approach that is based on designing and using compound heuristics called heuristics compositions. In fact, the developers have been using particular heuristics compositions since the end of the seventies as the approach that has not any alternative when one has to be concerned with complicated combinatorial problems for actual needs of real CAD/CAM systems in electronics industry. KAMBE et al.(1982) was one of the first papers which published a chain of sequential heuristics for problems of placing standard library cells and blocks of integrated circuits developed in CAD/CAM of the SHARP Corporation. It is difficult to draw a definite conclusion whether the concrete composition presented in the paper is practically used in real design process. Our doubt can be explained by the fact that such kinds of developments are very laborious and time-consuming, connected with numerous numeric experimentations on real size instances and usually constitute confidential "know-hows" of big companies. In MIATSELSKI (2009), we formulated the general principles of such kind of approach.

Heuristics compositions

Heuristics should be computationally efficient and tested experimentally in respect of solution quality. As a rule, an individual heuristic involving single evident, sometimes nadve, idea gives substitute of solution far from acceptable ones. However, this does not mean that such a heuristic should be irretrievably rejected. Experimentation with compositions of simple heuristics based on different ideas shows a significant growth of collective efficiency of such compositions in many cases. It can be also observed that the efficiency is positively influenced by the following factors:

1) engaging as more diverse ideas as possible by individual heuristics;

2) proper selection and arrangement of sequential heuristics for a composition to reach a proper trade-off between running time and solution quality;

3) determining proper numbers of iterations to be repeated or running time limits for the selected individual heuristics.

Realization of this principles is mainly based on computer experimentations with using databases of the test instances and benchmarks, public or private.

Usually (see for example REINELT 1994), two kinds of heuristics are distinguished. A heuristic of the first of kind, so-called a construction heuristic, is intended for obtaining a start approximation in seeking acceptable quality solutions. The first element of any heuristic composition is a constructive heuristic which constructs a feasible solution according to some construction rule. A heuristic of the second kind, so-called iterative or improvement heuristic is trying to improve current feasible solution using some rule of solution modification. Such modifications called moves are accepted or not according to another rule. Every attempt of modifying the current solution constitutes an individual iteration of an algorithm. Thus, an improvement heuristic is characterized by two rules: move rule and move acceptance rule. We conclude our description of simple and compound heuristics by the following formalism: a composition of heuristics is a chain

$$H_0{*}H_1{*}...{*}H_m,$$

where:

H_0 – a construction heuristic,
$H_1,..., H_m$ – are improvement heuristics (not necessarily all different),
$*$ – denotes the operation of consecutive running of algorithms.

Every improvement heuristic in the chain takes the result obtained by the predecessor as its input data. This result contains the recurrent approximation and possibly some additional information useful for the successors. Obviously, all the heuristics use the instance data of the problem which is being solved. In accordance with the efficiency factors stated above, a designer of a heuristics composition has to select and arrange the heuristics in the composition and also determine positive integers $k_1,...,k_m$ which are iteration numbers of the corresponding improvement heuristics. Another approach may require determining the corresponding time limits t_0, $t_1,...$, t_m for the heuristics.

In MIATSELSKI (2009) optimization problems related to designing heuristics compositions are proposed. In particular, the basic problem is reduced to the problem on paths in a weighted digraph that has maximum total length in sense of weights under a restricted number of arcs. This graph problem is

solvable in time $O(n^3 \log m)$, where n is the vertex number of the digraph, m is the limit on the number of arcs. Some kind of our graph problem was considered in CHRISTOFIDES (1985).

Permutations: notation, basic notions, properties

Permutations will be the basic combinatorial objects to deal with in this paper. That is why we have to specify our notation and make some assumptions about permutations.

Let X be a finite linear ordered set, $\prec$ be the order relationship on X, IXI denote the number of members of X. Mainly, we will deal with the following cases of set X: 1) $M = \{1,2,..., m\}$, $N = \{1, 2,..., n\}$; 2) $M \times N = \{(1, 2), (1, 3),..., (m, n)\}$; 3) $N^2 = \{(1, 2), (1, 3), ., (n, n)\}$, where the members of each set are given according to order $\prec$ on X.

We define permutation (substitution) p to be a bijection on X and use notation taking into account the order on X. For example in case 1), we write $p = \left[\frac{1, 2,...,m}{p_1, p_2,...,p_m}\right]$, but mainly we will write simply $p = (p_1, p_2,...,p_m)$ when no confusion can arise. The set of all permutations on X is denoted by $S[X]$, with using traditional notation S_m or S_n in case 1), S_{mn} and S_{n^2} for cases 2) and 3).

Cyclic structure $\Sigma(p)$ of permutation p is its important feature. We use symbols < > to be brackets for sub-cycles in the cyclic structure. We illustrate these notions by the examples as follows: $m = 8$, $p = \left[\frac{1,2,3,4,5,6,7,8}{7,2,4,5,8,3,1,6}\right] =$ $= (p_1,p_2,...,p_8) = (7, 2, 4, 5, 8, 3, 1, 6)$ $\Sigma(p) =$ <1,7> <2> <3, 4, 5, 8, 6>.

$q = \left[\frac{1,2,3,4,5,6,7,8}{3,1,5,2,7,8,6,4}\right] = (3, 1, 5, 2, 7, 8, 6, 4)$; $\Sigma(q) =$ < 1,3,5,7,6,8,4,2 >.

We would like to recall that permutations are bijections on set X and thus permutations being mappings can be composed as mappings. We accept the sequence of fulfilling the mappings on rule "from right to left" and use symbol * for composition of permutations. In the context of our examples, we have the following composition of permutations: $q*p = \left[\frac{1,2,3,4,5,6,7,8}{4,7,8,2,1,6,3,5}\right]$. It is well-known that $S[X]$ is a group, called *the symmetric group,* operation * is not commutative, thus $S[X]$ is not an Abelian group, I$S[X]$I = IXI!, for IS_mI = m!

We distinguish so-called cyclic permutations, i.e. such that their cyclic structure is reduced to one full cycle. We denote the set of all cyclic permutations by $C[X]$ or in case 1/ above C_m, C_n. Number of permutations I$C[X]$I = = (IXI-1)!, with IC_mI = (m-1)! for C_m.

Basic extreme permutation problems

Let $A = \|a_{ijkl}\|$ be a real $m \cdot m \cdot m \cdot m$ – four-index matrix, S_m denote the set of all permutations on symbols 1, 2, ..., m; $p = (p_1,p_2, ..., p_m) \in S_m$ is a permutation. Define

$$F(A; p) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} a_{ijp_ip_j} \tag{1}$$

General Quadratic Assignment Problem (GQAP) consists in seeking a permutation $p_0 \in S_m$ such that $F(p) \geq F(p_0)$ holds for every $p = (p_1,p_2, ..., p_m) \in S_m$. In other words, permutation p_0 provides the minimum value of objective function $F(A; p)$. In this meaning, permutation p_0 is xalled an optimal solution of the GQAP.

The GQAP is an intractable problem even for $m = 20–25$, with very special cases being *NP*-hard in strong meaning. For example, famous the Traveling Salesman Problem (TSP) is a very special case not only of the GQAP, but also of its special case well-known as the Koopmans-Beckman Problem (KBP). The Koopmans-Beckman Problem is defined as follows.

Let $C = \|c_{ij}\|$, $D = \|d_{ij}\|$, $E = \|e_{ij}\|$, be $m \cdot m$ – matrices. Find a permutation $p_0 \in S_m$ such that objective function

$$f(C, D, E; p) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} d_{ij}c_{p_ip_j} + \sum_{i=1}^{i=m} e_{ip_i} \tag{2}$$

attains its minimum $f(p_0)$ on S_m. It is easy to see that we obtain the TSP with matrix $C = \|c_{ij}\|$ if we set $e_{ij} = 0$ for all e_{ij} and let matrix $D = \|d_{ij}\|$ take the form:

$$\begin{bmatrix} 0 & 1... & 0 \\ \vdots & \ddots & 1 \\ 1 & ... & 0 \end{bmatrix}$$

In other words, D is the matrix representation of cyclic permutation $p = <1, 2, 3, ..., n>$ Certainly, another cyclic permutation can be selected instead of p.

Notice that matrix E is often zero-matrix and the second term in (2) is omitted, i.e. the KBP takes the followings form which is called simply QAP: find $p_0 \in S_m$ that gives the minimum of objective function

$$f(C, D; p) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} d_{ij} c_{p_i p_j} \quad (3)$$

on S_m

We would like to emphasize an important role of the QAP with objective (3) as one of the basic models used in different fields such as logistics, allocating production units and especially in computer science for goals of placing the library cells and blocks when integrated circuits are being designed, embedding graphs, forced network clustering, encoding (numerating) network's nodes. The two latter tasks are stated as follows.

I. Forced Network Clustering (cutting, decomposition)

Given a weighted graph $G = (V, E; w: E \to R^+)$, $|V| = m$, a partition $m = \sum_{k=1}^{k=s} m_k$ find a vertex-disjoint division of graph G into subgraphs $G_1, G_2, ..., G_s$ with vertex numbers respectively $m_1, m_2, ..., m_s$. The objective is to minimize the sum of edge weights w_{ij} of "cut" (external) edges, i.e. edges whose vertices belong to distinct subgraphs. This problem is reduced to the QAP with objective (3), where matrix C is the adjacency matrix of graph G, matrix $D = U\text{-}B$, where $U = \|u_{ij}\|$ is matrix with all its elements $u_{ij} = 1$. B = Diag $[B_1, B_2, ..., B_s]$ is a diagonal-block matrix, where blocks $B_1, B_2, ..., B_s$ are quadratic sub-matrices of sizes respectively $m_1, m_2, ..., m_s$ filled with units.

II. Encoding network's nodes

This problem consists in assigning numbers 1, 2, ..., m to vertices of graph $G = (V, E)$ in order to minimize sum $\sum_{i=1}^{i=m} \sum_{j=1}^{j=m} I\ p_i - p_j\ I$, where permutation $p = (p_1, p_2, ..., p_m)$ presents the assigned numbers. In the corresponding QAP, matrix C is the adjacency matrix of graph G. For matrix $D = \|d_{ij}\|$, $d_{ij} = \mathrm{I}i\text{-}j\mathrm{I}$.

Another version of the QAP, so-called the bottleneck QAP is obtained by substituting objective for objective function $f(C, D; p) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} d_{ij} c_{p_i p_j}$ for objective $g(C, D; p) = MIN_{ij=1,...,n} \{d_{ij} c_{p_i p_j}\}$. In the context of problems of encoding the graph vertices, we would like to emphasize that both the problems play an substantial role in operating on sparse matrices. The bottleneck QAP is actually another form of the minimum bandwidth problem for a sparse matrix.

Let us turn to the second term of (2), which actually is the value of objective in the following linear assignment problem (LAP). Given $m \cdot m$-matrix $E = \|e_{ij}\|$, *LAP* consists in seeking a permutation $p_0 \in S_m$ that minimizes objective function

$$V(E;\ p) = \sum_{i=1}^{i=m} e_{ip_i} \tag{4}$$

on S_m.

D.A. Suprunenko (see SUPRUNENKO, MIATSELSKI 1973) introduced conception of general linear assignment problem (GLAP) in the following way. Given matrix $E = \|e_{ij}\|$, and a nonempty subset $H \subseteq S_m$, find a permutation $p_0 \in H$ that minimizes function $V(E; p) = \sum_{j=1}^{j=m} e_{ip_i}$ on H. It can be easily seen that we obtain the traveling salesman problem when $H = C_m$. D.A. Suprunenko also introduced into consideration a special case of the GLAP, so- called problem of minimizing linear form on a permutation set $H \subseteq S_m$, in fact, problem of minimizing the scalar product (MSP) of two n-vectors on $H \subseteq S_n$. That means we have one more objective function defined for two vectors $a = (a_1, a_2, ..., a_m)$, $b = (b_1, b_2, ..., b_m)$:

$$v(a,\ b;\ p) = \sum_{i=1}^{i=m} a_i p_{p_i} \tag{5}$$

This case is distinguished by the following well-known smart fact. Without loss of generality we can assume that ordering $a_1 \le a_2 \le ... \le a_m$ holds. Then minimum (maximum) of $v(a, b; p) = \sum a_i b_{p_i}$ on S_m is attained on a permutation $p = (p_1, p_2, ..., p_m) \in S_m$ such that inequalities $b_{p1} \le b_{p2} \le ... \le b_{pm}$ ($b_{p1} \ge b_{p2} \ge ... \ge b_{pm}$ are satisfied.

Here we indicate a few relationships among problems and their objectives given above. First of all, notice that function (4) is evidently reduced to (5) when its matrix $E = \|e_{ij}\|$ is of rank 1. Really, rank 1 implies proportionality of each row of $E = \|e_{ij}\|$ to one of them, for example, to the first column. Then we select this column to be vector a and the sequence of proportionality coefficients to be vector b for the *MSP*. Moreover, as it is shown in (SUPRUNENKO, MIATSELSKI 1973) the following proposition takes place:

Let $m \cdot m$-matrix $E = \|e_{ij}\|$ of LAP have rank r, then there exist r pairs of m-vectors $a_{k,k=1,...r}$, $b_{k,k=1,...r}$, such that

$$V(E;\ p) = \sum_{k=1}^{k=r} v(a_k,\ b_k;\ p) \tag{6}$$

holds for any $p \in S_m$ and, conversely, $\sum_{k=1}^{k=s} v(a_k, b_k; p) = \mathrm{V}(E; p)$, where $E = \|e_{ij}\|$, $e_{ij} = \sum_{k=1}^{k=s} a_{ik}\, b_{jk}$ with rank of matrix E not exceeding number s.

Let us return to the GQAP with four-index matrix $A = \|a_{ijkl}\|$. We covert matrix A to its two-index representation $\tilde{A} = \| \tilde{a}_{(i,j)(kl)}\|$ which is a two-index $m^2 \cdot m^2$-matrix, with its elements indexed by ordered set $M^2 = \{(1, 2), (1, 3) ,..., (m,m)\}$ of pairs single indices.

Proposition 1 Let matrix $\tilde{A} = \| \tilde{a}_{(i,j)(kl)}\|$ obtained as result of converting matrix A of the GQAP to a two-index matrix have rank r, then there exist r pairs of matrices $C_{k,k=1,...,r}$, $D_{k,k=1,...,r}$ such that

$$F(A; p) = \sum_{k=1}^{k=r} f(C_k, D_k; p)$$

holds for each $p_0 \in S_m$ and, conversely, $\sum_{k=1}^{k=s} f(C_k, D_k; p) = F(A; p)$, where rank r, of $\tilde{A}$ does not exceed value s.

Proposition 1 can be derived from the previous proposition after some preparations including re-indexation of set $M^2 = \{(1, 2), (1, 3), ..., (m, m)\}$ changing pair indices for single ones.

Operations over permutation sets: some algebraic structures and combinatory constructions

Our definitions of operations over permutation sets will be based on the ways to construct permutations on sets $M + N = \{1, 2, ..., m, m + 1, m + 2, ..., m + n\}$, $M \cdot N = \{(1, 2), (1, 3), ..., (m, n)\}$ and $M^2 = \{(1, 2), (1, 3), ..., (m, m)\}$ ("big permutations") by means of permutations of S_m or S_n ("building blocks", "small permutations"). These definitions come from some concepts which had been introduced under different names for permutation groups, see for example (SUPRUNENKO 1996). We present these notions defined for arbitrary permutation sets out of group theory context.

Name: Direct sum; Operation symbol: $\oplus$; Constructing direct sum of single permutations: $p = (p_1, p_2, ..., p_m) \in S_m$, $q = (q_1, q_2, ..., q_n) \in S_n \overset{\text{reslt}}{\Longrightarrow}$ [Action on united chain $M + N$, $p \oplus q$:$i \overset{\text{def}}{\longrightarrow} p_i$ for i, $1 \le i \le$ m, $p \oplus q$: $j \overset{\text{def}}{\longrightarrow} q_j$ for j, $m + 1 \le j \le m + n$;] m + n;]

Constructing direct sum of sets $G \subseteq S_m$, $H \subseteq S_n$: $G \oplus H \overset{\text{def}}{=} \{(p, q): p \in G, q \in H\}$. Comment: Direct sum of permutation sets $G \oplus H$ acts independently on disjoint parts of chain $M+N=\{1, 2, ..., m, m+1, m+2, ..., m+ n\}$, I$G \oplus H$I = I$G$I I$H$I, in particular I$S_m \oplus S_n$I = m!n!

Name: Direct product; Operation symbol: $\oplus$; Constructing direct product of single permutations: $p = (p_1, p_2, ..., p_m) \in S_m$, $q = (q_1, q_2, ..., q_n) \in S_n \overset{\text{reslt}}{\Longrightarrow}$ [Action on set $M+N$, $p \oplus q$:$(i, j) \overset{\text{def}}{\longrightarrow} (p_i, q_j)$ for i, $1 \le i \le m$, j, $1 \le j \le n$;]

Constructing direct product of sets $G \subseteq S_m$, $H \subseteq S_n$: $G \oplus H \stackrel{\text{def}}{=} \{(p, q): p \in G, q \in H\}$. Comment: $\mathrm{I}G \oplus H\mathrm{I} = \mathrm{I}G\mathrm{I}\ \mathrm{I}H\mathrm{I}$, in particular $\mathrm{I}S_m \oplus S_n\mathrm{I} = m!n!$

Name: Wreath product; Operation symbol: ≀ Constructing wreath product of single permutations: $p = (p_1, p_2, ..., p_m) \in S_m$, $q^1 = (q_1^1\, q_2^1, ..., q_n^1) \in S_n$, ..., $q^m = (q_1^m q_2^m, ..., q_n^m) \in S_n \stackrel{\text{resit}}{\Longrightarrow}$ [Action on set $M \cdot N$; $p \wr q$: $(i, j) \stackrel{\text{def}}{\longrightarrow} (p_i, q_j^i)$ for i, $1 \le i \le m$, j, $1 \le j \le n$;]

Construction of wreath product of sets $G \subseteq S_m$, $H \subseteq S_n$: $G \wr H \stackrel{\text{def}}{=} \{(p; q^1, q^2, ..., q^m): p \in G, q^i \in H, 1 \le i \le m\}$. Comment: $\mathrm{I}G \wr H\mathrm{I} = \mathrm{I}G\mathrm{I}\ \mathrm{I}H\mathrm{I}^{\mathrm{IGI}}$, in particular $\mathrm{I}S_m \wr S_n\mathrm{I} = m!n!^m$.

Name: Diagonal of $S_m \otimes S_m$; Let $m=n$ hold. Notation and definition: diag $[S_m \otimes S_m] \stackrel{\text{def}}{=} \{(p, q): p \in S_n, q \in S_n, p = q\}$.

The following relationships for sets $G \in S_m$, $H \in S_n$ can be obtained on the base of constructions of permutation sets described above.

$$G \otimes H \subset G \wr H \subset S_{mn},\ (m > 1,\ n > 1) \tag{7a}$$

$$\text{diag}\,[S_m \otimes S_m] \subset S_m \otimes S_m \subset S_m \wr S_m \subset S_{m^2},\ m > 1 \tag{7b}$$

These inclusions enable us to detect relationships among the permutation problems described above. We will also use abbreviation $\overline{S_m}$ for diag $[S_m \otimes S_m]$ First of all, we will consider the most general case of matrices, namely our input data are given by four-index matrix $A = \|a_{ijkl}\|$ i.e. we will be concerned with the proper (initial) GQAP which corresponds to position of diag $[S_m \otimes S_m]$ in the inclusions chain (7*b*). At this point, we step up in our generalizations and extend the set of feasible permutations of the GQAP permitting the permutations to belong to sets wider than diag $[S_m \otimes S_m]$, in particular to the successive terms of chain (7*b*). In this respect, we follow the concept of the GLAP above, with transferring the idea of the LAP to the GQAP. Thus, the next term is $S_m \otimes S_m$ or more general $S_m \otimes S_n$.

The General QAP on $S_m \otimes S_m$ has all indications to be intractable in both theoretical and practical meaning. Therefore, we restrict ourselves by the case when matrix $\tilde{A} = \|\tilde{a}_{(i,j)(kl)}\|$ obtained in result of converting matrix $A = \|a_{ijkl}\|$ of the GQAP to a two-index matrix has rank 1. It follows from proposition 1 that $F(A; p) = f(C_1, D_1; p)$ or simply $F(A; p) = f(C, D; p)$, where according to our conventions above we deals with objective function $f(C, D; p) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} d_{ij} c_{p_i p_j}$ or more generally $f(C, D; p,q) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=n} d_{ij} c_{p_i q_j}$ on $S_m \otimes S_m$ or $S_m \otimes S_n$ respectively. Unlike the QAP the problem of minimization $f(C, D; p) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=n} d_{ij} c_{p_i q_j}$ on $S_m \otimes S_n$ so far has attracted insufficient attention of the researches. Moreover, there is not any generally accepted name for the problem. Nevertheless, the

problem has practical applications similar to those of the QAP, however with explicit accent on bipartite weighted graphs (networks).

Let $\bar{p} \in S_m$ be an arbitrary fixed permutation. Then we have $f(C, D; \bar{p}, q) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=n} d_{ij} c_{\bar{p}_i q_j} = \sum_{i=1}^{j=n} \sum_{j=1}^{i=m} d_{ij} c_{\bar{p}_i q_j} = \sum_{j=1}^{j=n} e_{jq_j}$, where $e_{jq_j} = \sum_{i=1}^{i=m} d_{ij} c_{\bar{p}_i q_j} = V(E_{\bar{p}}; q) = \sum_{j=n} e_{jq_j}$ and $E_{\bar{p}} = \|e_{jk}\|$ is an $n \cdot n$-matrix.

Thus, when one of the permutations in the objective function $f(C, D; \bar{p}, q)$ is fixed we actually deal with the LAP. In respect of computational complexity, there is a principal difference between the QAP and the LAP: the first is NP- hard while the other is efficiently solvable in time $O(n^3)$. For this reason, it seems to be promising for solving the QAP to come down it to a finite not too numerous series of LAPs. The following algorithm exploits this approach.

Algorithm Min-on-direct-product

Input: $C = \|c_{ij}\|$, $D = \|d_{ij}\|$ /* $n \cdot n$-matrices;

Output: $p_0 \in S_m$, $q_0 \in S_n$ /*a pair of permutations which is accepted to be an approximate solution to the problem of minimizing objective function $f(C, D; p, q) = \sum_{i=1}^{i=m} \sum_{j=1}^{j=n} d_{ij} c_{p_i q_j}$ on $S_m \otimes S_n$*/;

Main Program:

1) Select: $x \in S_m$; $y \in S_n$; /* Selection of initial permutations;

Do

{2) find $y^\wedge$ that minimize $f(C, D; x, q)$ on $\{x\} \otimes S_n$ while x is fixed; /*in fact, solving a LAP;

2a) $y := y^\wedge$;

3) find $x^\wedge$ that minimize $f(C, D; p, y)$ on $S_m \otimes \{y\}$ while y is fixed; /*in fact, solving a LAP;

3a) $x := x^\wedge$;}

Until $x^\wedge = x$ & $y^\wedge = y^\wedge$;

4) $p_0 := x^\wedge$; $q_0 := y^\wedge$; /* The result has been obtained, further improvement of the current solution in this way is impossible; */

End of Algorithm

This algorithm may be used to be a part of heuristics compositions, both as a construction heuristic and improving one. Obviously, algorithm Min-on--direct-product can be modified by introduction of random repeated choices $x \in S_m$; $y \in S_n$ to instruction 1) intended for selection of initial permutations. The final solution is the best of (p_0, q_0) obtained in result of many trials $x \in S_m$; $y \in S_n$.

Conclusions and future work

In this paper, we have concerned a unified approach to compound heuristics and uniform presentation of the basic extreme permutations problems as well as related structures. This constitutes a basis for further research in connection with the general assignment problem on the wreath product in inclusion chain $(7a, b)$ and using metrics on the symmetric group for seeking heuristic solutions.

References

BURKARD R.E., CELA E., DEMIDENKO V.M., MIATSELSKI M, WOEGINGER G.J. 1995. *A unified approach to simple special cases of extremal permutation problems.* Optimierung und Kontrolle. K-FU Graz, TUGraz, p 12.

BURKARD R.E., CELA E., DEMIDENKO V.M., MIATSELSKI M., WOEGINGER G.J. 1998. A *unified approach to simple special cases of extreme permutation problems.* Optimization, 44: 123–138.

CELA E. 1998 *The quadratic assignment problem: Theory and algorithms*. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 286.

CHRISTOFIDES N. 1985. *Graph theory: an algorithmic approach.* Academic Press, p 428.

KAMBE T., KIMURA S., INUFUSHI T., OKUDA N., NISHIOKA I. 1982. *A Placement Algorithm for Polycell LSI and its Evaluation.* Design Automation, 19th Conference, Las Vegas, p. 123–132.

LAWLER E.L., LENSTRA J.K., RINNOOY KAN A.H.G., SHMOYS D.B. 1985. *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization.* Wiley.

MIATSELSKI M. 2009. *Compound heuristics as means for practical solving theoretically hard problems coming from computer-aided design.* Annals of Higher School of Computer Science and Economics, Olsztyn, p. 37–46.

REINELT G. 1994. *The traveling salesman problem.* Lecture Notes in Computer Science, 840: 324.

SUPRUNENKO D.A. 1996. *Permutations groups.* Nauka i Technika, p. 366.

SUPRUNENKO D.A., MIATSELSKI M.M. 1975. *The assignment problem and minimization of the sum of linear forms on the symmetric group.* Cybernetics and Systems, 3: 64–68.

Technical Sciences, 2018, 21(1), 49–63

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Quarterly peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: www.uwm.edu.pl/techsci/

ANALYSIS OF VARIATIONS IN AND CORRELATIONS BETWEEN SELECTED PHYSICAL PARAMETERS OF COMMON BEECH (*FAGUS SILVATICA L.*) NUTS

***Zdzisław Kaliniewicz*[1], *Piotr Markowski*[1], *Andrzej Anders*[1], *Paweł Tylek*[2], *Zbigniew Krzysiak*[3], *Wojciech Wasielewski*[1]**

[1] Department of Heavy Duty Machines and Research Methodology
University of Warmia and Mazury in Olsztyn
[2] Department of Forest Work Mechanization
University of Agriculture in Kraków
[3] Department of Mechanical Engineering and Automatic Control Engineering
University of Life Sciences in Lublin

Received 10 April 2016, accepted 30 October 2017, available online 6 November 2017.

K e y w o r d s: geometric parameters, mass, aspect ratio, separation.

A b s t r a c t

The basic dimensions and the mass of common beech nuts and seeds from five nut batches, harvested from tree stands in northern Poland, were determined. Environmental conditions had a greater influence on seed plumpness than the age of tree stands. The results of measurements were analyzed statistically by analysis of variance, correlation analysis and linear regression analysis. Despite differences in their plumpness, nuts were characterized by nearly identical cross-sections which resembled an equilateral triangle. The thickness of nuts and seeds was highly correlated with their mass, and this information can facilitate seed husking and separation into mass categories. Before and after husking, seeds should be separated with the use of a mesh screen with longitudinal openings. Medium-sized (most numerous) seeds were separated into the following plumpness categories using a screen separator with ≠6 mm and ≠7 mm openings: 84% of moderately plump seeds, 3% of seeds with reduced plumpness, and 13% of plump seeds.

Symbols

m – seed mass [mg],
M – nut mass [mg],
SD – standard deviation of trait,

Correspondence: Zdzisław Kaliniewicz, Katedra Maszyn Roboczych i Metodologii Badań, Uniwersytet Warmińsko-Mazurski w Olsztynie, ul. Oczapowskiego 11/B112, 10-719 Olsztyn, phone: 48 89 523 39 34, e-mail: zdzislaw.kaliniewicz@uwm.edu.pl

t, w, l – seed thickness, width and length [mm],
T, W, L – nut thickness, width and length [mm],
w_1, w_2, w_3 – seed width, measured on three sides [mm],
W_1, W_2, W_3 – nut width, measured on three sides [mm],
V_s – coefficient of trait variation [%],
x – average value of trait,
x_{max}, x_{min} – maximum and minimum value of trait,
γ_1, γ_2, γ_3 – vertical angles of seed cross-section [°],
Γ_1, Γ_2, Γ_3 – vertical angles of nut cross-section [°].

Introduction

The common beech (*Fagus sylvatica* L.) is a deciduous tree species found in regions with a temperate marine climate. Its geographic range covers nearly all of Central and Western Europe (DITTMAR et al. 2003, VON OHEIMB et al. 2005, BOLTE et al. 2007, BRUS et al. 2012, JAWORSKI 2011, MAGRI 2008, PUKACKA, RATAJCZAK 2005). The common beech is a relatively slow-growing species, and it begins to produce seeds (nuts) at the age of 40–50 years in open stands and even later in dense stands. This monoecious species produces unisex and anemophilous flowers. Beech nuts mature in September to October to release 2, 3 or, less frequently, 4 triangular seeds. The external pericarp is brown, glossy, rather thin, flexible and resistant to crushing (Fig. 1). It contains one (most frequently) or two seeds (SUSZKA et al. 2000, JAWORSKI 2011).

Beech seeds are a delicacy for many forest animals, including wild boars, deer, squirrels, mice, jays and nutcrackers (SKRZYDŁOWSKI, PIOTROWSKI 2003, RUSCOE et al. 2005, JAWORSKI 2011). Seeds contain 30–36% fat, 25% protein, saponins, malic acid, citric acid and vanillic acid, 6% minerals, sugars and starch (REYES et al. 2006, PUKACKA, RATAJCZAK 2014). Raw seeds should not be consumed in large quantities because they contain small amounts of trimethylamine (fagine), a poisonous and hallucinogenic substance. Trimethylamine is removed by roasting, and purified seeds can be added to pastry and desserts or used as a coffee substitute. Beech seed oil has a very long shelf life, and its properties improve with storage time (Institute of Dendrology in Kórnik, www.idpan.poznan.pl).

Every 5 or 10 years, the yield of beech nuts per 1 ha reaches 4 tons, and seeds are produced by trees as old as 200 years (SUSZKA et al. 2000, ÖVERGAARD et al. 2007). Due to their abundance, beech nuts can be used widely in food production. According to BODYŁ and SUŁKOWSKA (2007), the demand for common beech seeds in Poland is met when approximately 10% of trees listed in the National Register of Forest Reproductive Material produce seeds. When other beech stands and seed consumption by animals are taken into account, the supply of beech seeds exceeds current demand, which indicates that the surplus could be used in food processing.

Fig. 1. View of common beech nuts: *a* – nut (seed with pericarp), *b* – husked seed with seed coat, *c* – husked seed without seed coat

The objective of this study was to determine mutual correlations between the geometric parameters and the weight of common beech nuts and seeds so as to maximize the efficiency of nut separation and husking processes.

Materials and Methods

The experimental material comprised five batches of common beech nuts supplied by a seed extraction plant in Jedwabno. Three batches were harvested from variously aged tree stands in one forest region, and two batches were obtained from similarly aged tree stands in other forest regions of northern Poland (Fig. 2). The analyzed batches were harvested from the following tree stands:

a) registration No. MP/1/1995/05, category of seed propagation material – from an identified source, region of origin – 103, municipality – Tolkmicko, geographic location – 54.18°N, 19.31°E, forest habitat – fresh forest, age – 105 years (symbol: CB-1);

b) registration No. MP/1/48559/08, category of seed propagation material – from an identified source, region of origin – 251, municipality – Kolno, geographic location – 53.54°N, 20.53°E, forest habitat – fresh forest, age – 129 years (symbol: CB-2a);

c) registration No. MP/1/12857/05, category of seed propagation material – from an identified source, region of origin – 103, municipality – Młynary, geographic location – 54.14°N, 19.40°E, forest habitat – fresh forest, age – 130 years (symbol: CB-2b);

d) registration No. MP/1/43920/05, category of seed propagation material – from an identified source (removed from the list), region of origin – 451, municipality – Lidzbark, geographic location – 53.16°N, 19.42°E, forest habitat – fresh forest, age – 124 years (symbol: CB-2c);

e) registration No. MP/1/10482/05, category of seed propagation material – from an identified source, region of origin – 103, municipality – Godkowo, geographic location – 54.06°N, 19.54°E, forest habitat – fresh forest, age – 155 years (symbol: CB-3).

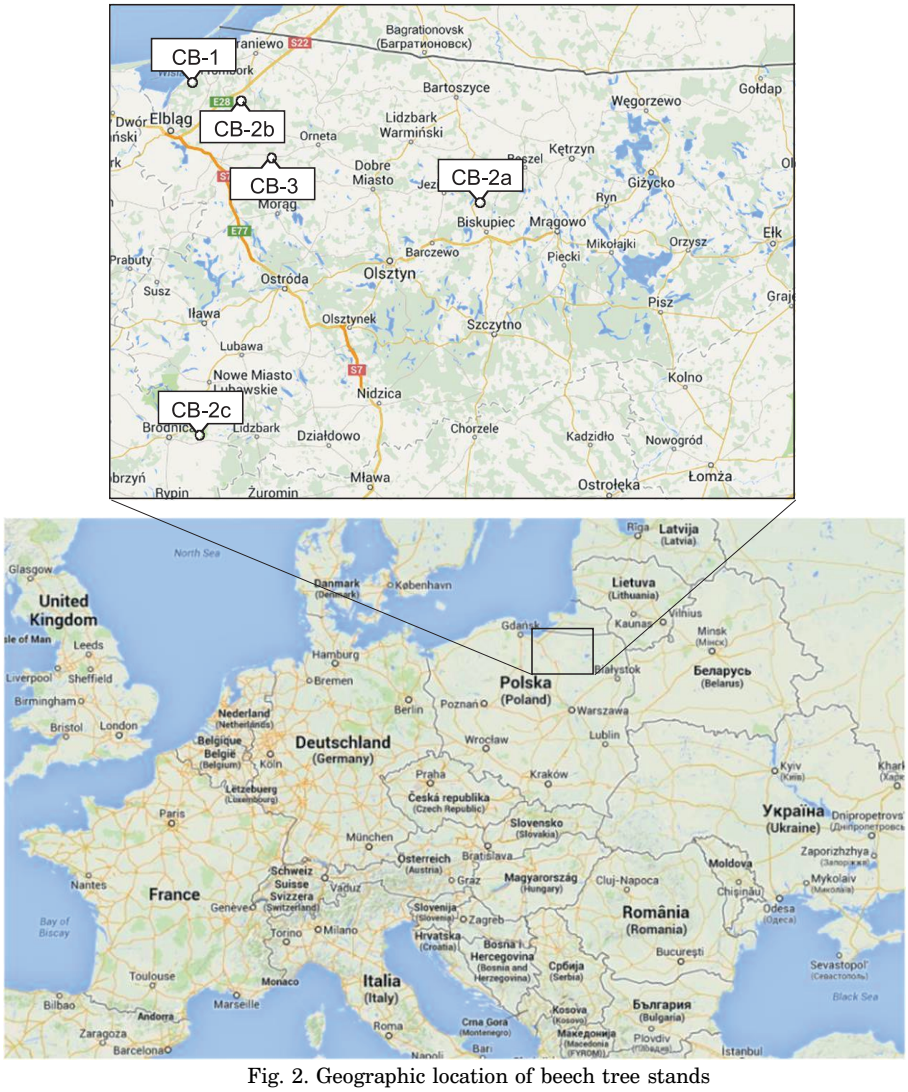

Fig. 2. Geographic location of beech tree stands

Analytical samples (initial samples had the weight of 5 kg) from every batch of nuts were divided by halving (*Nasiennictwo*... 1995). Initial samples were halved, and one half was randomly selected for successive halving. The above procedure was repeated to produce samples of around 100 nuts each.

The length and width of beech nuts were determined with the use of the MWM 2325 laboratory microscope (PZO Warszawa, Poland) to the nearest 0.02 mm (one measurement covered two micrometer readings), and nut thickness was determined with a dial thickness gauge to the nearest 0.01 mm. Beech nuts are triangular, and their width and thickness were measured on each side. Nut weight was determined on the WAA 100/C/2 laboratory scale (RADWAG Radom, Polska) to the nearest 0.1 mg. All measurements were performed according to the methods previously described by KALINIEWICZ et al. (2011) and KALINIEWICZ and POZNAŃSKI (2013).

Beech nuts were husked manually to extract seeds. The seeds were subjected to the above measurements, and correlations were determined between nut and seed parameters. The symbols describing the analyzed parameters were given in uppercase letters for nuts and in lowercase letters for seeds. Nutlets containing two seeds were excluded from further analysis due to significant variations in their shape and dimensions relative to typical seeds. Such nutlets accounted for less than 2% of the evaluated samples, which is consistent with the results reported by KALINIEWICZ et al. (2015). The analyzed nut samples had the following size: CB-1 – 105, CB-2a – 115, CB-2b – 108, CB-2c – 103, CB-3 – 109. In the examined samples, standard error for the evaluated physical parameters did not exceed:

– length of nuts and seeds – 0.3 mm,
– width and thickness of nuts and seeds – 0.2 mm,
– mass of nuts and seeds – 13 mg.

Vertical angles of nut and seed cross-sections (at the widest point) were determined with the use of the cosine formula (Carnot's theorem) which was transformed as follows:

$$\Gamma = \arccos \frac{W_2^2 + W_3^2 - W_1^2}{2 \cdot W_2 \cdot W_3} \tag{1}$$

Husked seeds were divided into three plumpness categories: seeds with reduced plumpness ($m < 150$ mg), moderately plump seeds ($m = 150 \div 250$ mg) and plump seeds ($m > 250$ mg).

The results of measurements and calculations were processed in Statistica v. 10 based on standard statistical procedures, including one-way ANOVA, correlation analysis and linear regression analysis (RABIEJ 2012). The results were regarded as statistically significant at $p = 0.05$.

Results

The physical parameters of common beech nuts are presented in Table 1. The least developed nuts (lowest average length, width, thickness and mass) were noted in batch CB-2c. Batch CB-2a contained the plumpest nuts. Nuts harvested in the same region (batches CB-1, CB-2b and CB-3) differed mostly in length (differences in nut thickness and width between batches were not statistically significant). A comparison of the physical parameters of nuts harvested from similarly aged tree stands (batches CB-2a, CB-2b and CB-2c) revealed more significant differences. The above implies that the evaluated traits of beech nuts can also be influenced by the local climate.

Differences nut width, measured on each of the three sides and arranged in descending order, are presented in Table 2. Statistically significant differences in average width and, consequently, in vertical angles of nut cross-sections

Table 1

Physical parameters of common beech nuts from five batches

Parameter	Batch				
	CB-1 $x \pm SD$	CB-2a $x \pm SD$	CB-2b $x \pm SD$	CB-2c $x \pm SD$	CB-3 $x \pm SD$
L [mm]	16.86 ± 1.43^{C}	18.16 ± 1.47^{a}	17.08 ± 1.42^{Bb}	16.13 ± 1.37^{c}	17.46 ± 1.35^{A}
W [mm]	9.21 ± 0.99^{A}	9.58 ± 1.19^{a}	9.18 ± 1.00^{Ab}	8.75 ± 1.07^{c}	9.12 ± 1.03^{A}
T [mm]	8.03 ± 0.84^{A}	8.31 ± 0.98^{a}	8.00 ± 0.83^{Ab}	7.61 ± 0.88^{c}	7.93 ± 0.85^{A}
M [mg]	294.20 ± 62.56^{B}	314.71 ± 61.64^{a}	311.52 ± 60.36^{Aa}	275.93 ± 61.38^{b}	301.49 ± 67.97^{B}

$^{A, B, C}$ – different letters indicate statistically significant differences in a given parameter between nuts harvested from the same region.
$^{a, b, c}$ – different letters indicate statistically significant differences in a given parameter between nuts harvested from similarly aged tree stands.

Table 2

Width of beech nuts, measured on all sides, and vertical angles of nut cross-sections in five batches of common beech nuts

Parameter	Batch				
	CB-1 $x \pm SD$	CB-2a $x \pm SD$	CB-2b $x \pm SD$	CB-2c $x \pm SD$	CB-3 $x \pm SD$
W_1 [mm]	9.90 ± 0.86^{Ba}	10.42 ± 1.16^{Aa}	9.89 ± 0.90^{Ba}	9.44 ± 1.04^{Ca}	9.85 ± 0.97^{Ba}
W_2 [mm]	9.04 ± 0.87^{Bb}	9.34 ± 0.95^{ab}	9.06 ± 0.84^{Bb}	8.57 ± 0.90^{Cb}	8.94 ± 0.84^{Bb}
W_3 [mm]	8.70 ± 0.84^{Bc}	8.97 ± 0.94^{Ac}	8.60 ± 0.80^{Bc}	8.24 ± 0.88^{Cc}	8.56 ± 0.82^{Bc}
Γ_1 [°]	68.01 ± 4.19^{Ba}	69.51 ± 5.56^{Aa}	68.19 ± 4.38^{ABa}	68.36 ± 4.12^{ABa}	68.62 ± 4.60^{ABa}
Γ_2 [°]	57.64 ± 2.55^{ABb}	56.89 ± 3.43^{Bb}	58.08 ± 2.61^{Ab}	57.44 ± 2.42^{ABb}	57.51 ± 2.81^{ABb}
$\Gamma 3$ [°]	54.35 ± 2.27^{Ac}	53.60 ± 3.47^{Ac}	53.73 ± 2.70^{ac}	54.17 ± 2.47^{Ac}	53.87 ± 2.86^{Ac}

$^{A, B, C}$ – different letters indicate statistically significant differences in a given parameter between nuts from different batches.
$^{a, b, c}$ – different letters indicate statistically significant differences in a given parameter between nuts from the same batch.

were observed in all batches. A comparison of the widest, average and narrowest nut sides in the analyzed batches indicates that CB-2a and CB-2c differed significantly from the remaining batches, whereas no significant differences were noted in their respective vertical angles. Statistically significant differences were observed only in angle Γ_1 between batches CB-2a and CB-1 and in angle Γ_2 between batches CB-2a and CB-2b. The above results indicate that cross-sectional shape is not determined by harvesting site or the age of trees.

The correlations between the physical parameters of seeds were generally similar to those noted between the analyzed parameters of nuts (Fig. 3). Significant differences were observed only in batch CB-2c which was characterized by the smallest seed dimensions and the lowest seed mass. The average values of seed properties varied in the following range: length – from 12.1 to 13.2 mm, width – from 6.8 to 7.3 mm, thickness – from 6.1 to 6.5 mm, mass

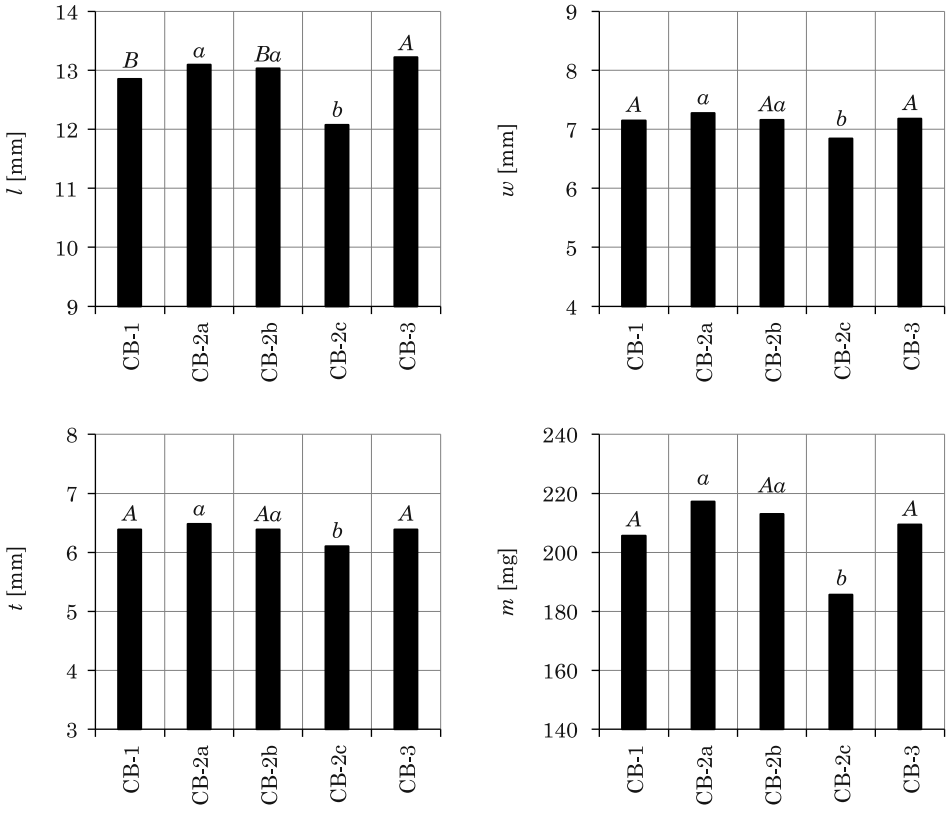

Fig. 3. Average values of seed parameters: *A*, *B* – different letters indicate statistically significant differences in a given parameter between nuts from the same region; *a*, *b* – different letters indicate statistically significant differences in a given parameter between nuts from similarly aged tree stands

– from 185.7 to 217.2 mg. Similarly to nuts, more statistically significant differences in physical parameters were noted between seeds harvested from similarly aged tree stands in different regions than between seeds obtained from variously aged tree stands in the same region.

The width of seeds (extracted from nuts), measured on each side and at the widest point, from every analyzed batch is presented in Table 3. A comparison of the width of corresponding seed sides did not reveal statistically significant differences, except for batch CB-2c. Significant differences in vertical angles were not observed in any of the evaluated batches.

Table 3

Width of beech seeds, measured on all sides, and vertical angles of seed cross-sections in five batches of common beech nuts

Parameter	Batch				
	CB-1 $x \pm SD$	CB-2a $x \pm SD$	CB-2b $x \pm SD$	CB-2c $x \pm SD$	CB-3 $x \pm SD$
w_1 [mm]	7.64 ± 0.67^{Aa}	7.81 ± 0.70^{Aa}	7.63 ± 0.65^{Aa}	7.31 ± 0.79^{Ba}	7.69 ± 0.78^{Aa}
w_2 [mm]	7.05 ± 0.68^{Ab}	7.17 ± 0.67^{ab}	7.10 ± 0.63^{Ab}	6.76 ± 0.74^{Bb}	7.07 ± 0.69^{Ab}
w_3 [mm]	6.76 ± 0.67^{Ac}	6.86 ± 0.65^{Ac}	6.76 ± 0.60^{Ac}	6.47 ± 0.70^{Bc}	6.78 ± 0.68^{Ac}
γ_1 [°]	67.25 ± 3.43^{Aa}	67.75 ± 4.55^{Aa}	66.86 ± 3.61^{Aa}	67.06 ± 3.62^{Aa}	67.45 ± 4.23^{Aa}
γ_2 [°]	58.20 ± 2.25^{Ab}	58.01 ± 3.05^{Ab}	58.70 ± 2.24^{Ab}	58.34 ± 2.54^{Ab}	58.08 ± 2.65^{Ab}
γ_3 [°]	54.55 ± 2.05^{Ac}	54.24 ± 2.95^{Ac}	54.44 ± 2.51^{Ac}	54.60 ± 2.49^{Ac}	54.47 ± 2.78^{Ac}

$^{A, B, C}$ – different letters indicate statistically significant differences in a given parameter between seeds from different batches.

$^{a, b, c}$ – different letters indicate statistically significant differences in a given parameter between seeds from the same batch.

Despite local statistical differences, seeds from the five tested batches were regarded as homogeneous. The following ratios were determined between the compared physical parameters of nuts and seeds: $L/l = 1.34 \pm 0.06$, $W/w = 1.29 \pm 0.07$, $T/t = 1.26 \pm 0.07$, $M/m = 1.47 \pm 0.09$.

A linear correlation analysis revealed that all physical parameters of common beech nuts and seeds (Tab. 4) were significantly correlated at 0.05. The highest value of the correlation coefficient (0.98) was observed between the masses of nuts and seeds, and the lowest value (0.29) – between seed thickness and nut length. Strong correlations were generally noted between the corresponding parameters of nuts and seeds (from 0.87 for length to 0.98 for mass), which influenced the value of the coefficient of determination in linear regression equations describing the above traits.

Equations where the coefficient of determination is higher than 0.2 are presented in Table 5. Seed mass was significantly correlated with all evaluated traits of beech nuts and seeds. The highest value of the coefficient of determination and the equation with the highest proportion of explained variation

Table 4
Coefficients of linear correlation between selected physical parameters of beech nuts and seeds

Parameter	W	T	M	l	w	t	m
L	0.338	0.338	0.557	0.869	0.303	0.290	0.532
W	1	0.413	0.666	0.327	0.894	0.343	0.610
T		1	0.701	0.336	0.385	0.897	0.646
M			1	0.644	0.749	0.762	0.979
l				1	0.359	0.350	0.644
w					1	0.442	0.736
t						1	0.752

All correlations are statistically significant at 0.05.

were noted for the correlation between seed mass and nut mass (0.96). Nut thickness and seed thickness were most highly correlated with seed mass. The above observation suggests that common beech seeds would be most effectively separated into fractions with the use of a mesh screen with longitudinal openings.

Table 5
Regression equations for the physical parameters of beech seeds

Equation	Coefficient of determination R_2	Standard error of estimate
$l = 0.672\ L + 1.344$	0.755	0.597
$l = 0.012\ M + 9.241$	0.415	0.922
$l = 0.016\ m + 9.663$	0.415	0.922
$w = 0.649\ W + 1.173$	0.801	0.353
$w = 0.009\ M + 4.362$	0.560	0.525
$w = 0.012\ m + 4.726$	0.541	0.537
$t = 0.700\ T + 0.761$	0.804	0.313
$t = 0.008\ M + 3.833$	0.581	0.458
$t = 0.011\ m + 4.156$	0.565	0.467
$m = 17.056\ L - 86.177$	0.283	42.419
$m = 27.972\ W - 50.152$	0.372	39.683
$m = 35.673\ T - 78.278$	0.417	38.232
$m = 0.763\ M - 22.371$	0.959	10.110
$m = 26.753\ l - 137.726$	0.415	38.301
$m = 46.504\ w - 125.002$	0.541	33.929
$m = 53.191\ t - 131.255$	0.565	33.022

In line with the adopted classification system, husked nutlets will produce 13% of poorly filled seeds (m < 150 mg), 69% of moderately plump seeds (m = 150÷250 mg) and 18% plump seeds (m > 250 mg). Nuts can be separated into six fractions based on their thickness (Figure 4). Every fraction contains nuts whose seeds can be classified into three mass categories. The only exceptions are the finest (T < 6.01 mm) and coarsest (T > 10 mm) fractions which do not contain plump seeds and poorly filled seeds, respectively.

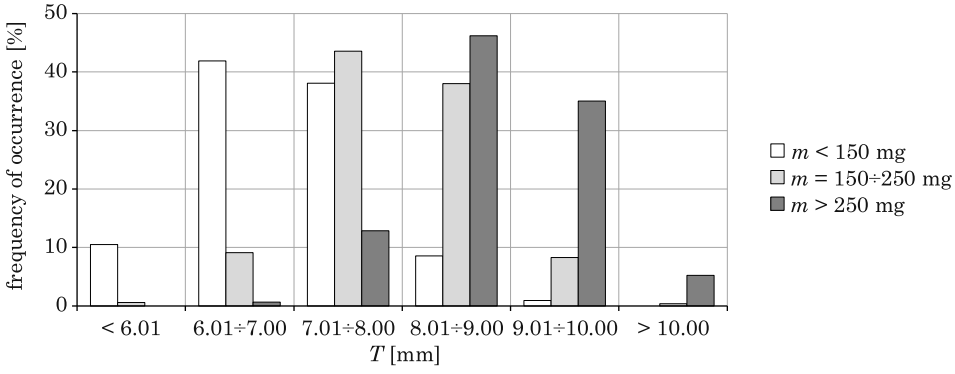

Fig. 4. Distribution of nut thickness across three mass categories of husked seeds

The data presented in Table 4 indicates that seed mass is most highly correlated with seed thickness, which implies that seeds can be effectively sorted with the use of a mesh screen with longitudinal openings. Seeds can be sorted into five size fractions (Fig. 5) with evenly distributed mass. Two mesh screens with ≠6 mm and ≠7 mm openings can be used to separate seeds into three size fractions, where the finest fraction contains approximately 89% of poorly filled seeds, 27% of moderately plump seeds and only 2% of plump seeds. The coarsest fraction will comprise only moderately plump seeds (approximately 10%) and plump seeds (approximately 58%). The above indicates that the medium-sized fraction will contain approximately 11% of poorly filled seeds, 62% of moderately plump seeds and 40% of plump seeds. The analyzed size fractions were characterized by the following share of plumpness categories:

– finest fraction ($t \leq 6$ mm) – 37.4% of poorly filled seeds, 61.4% of moderately plump seeds and 1.2% of plump seeds,

– medium fraction ($t = 6 \div 7$ mm) – 2.8% of poorly filled seeds, 83.9% of moderately plump seeds and 13.3% of plump seeds,

– coarsest fraction ($t > 7$ mm) – 41.3% of moderately plump seeds and 58.7% of plump seeds.

Discussion

The evaluated beech nuts were more filled than those analyzed by BODYŁ and SUŁKOWSKA (2007), and they were nearly identical with those examined by TYLEK (2010) and KALINIEWICZ et al. (2015). The average length of the evaluated beech nuts was similar to that of pumpkin seeds (JOSHI et al. 1993), squash seeds (JACOBO-VALENZUELA et al. 2011) and kidney beans (ALTUNTAS, DEMIRTOLA 2007).

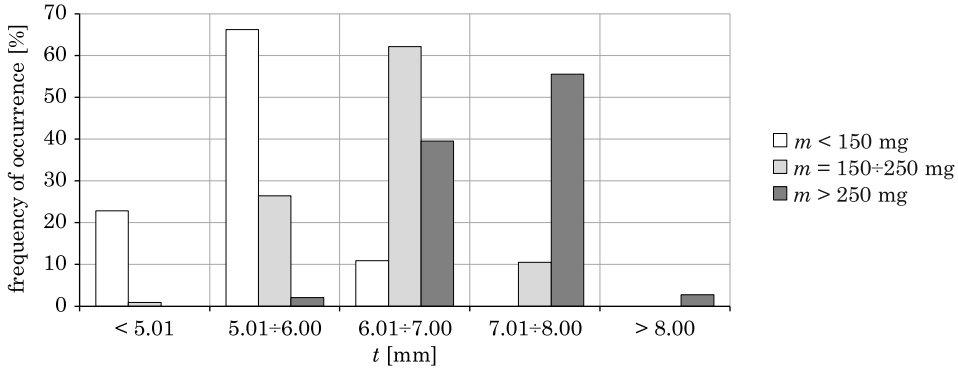

Fig. 5. Distribution of seed thickness across three weight categories

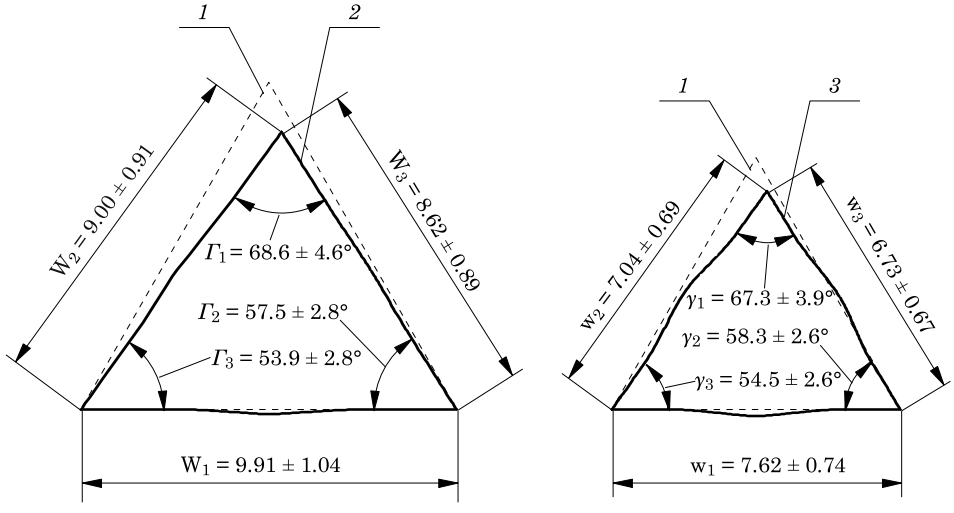

Fig. 6. Cross-sectional shape of the largest beech nuts (*2*) and seeds (*3*) shown against an equilateral triangle (*1*)

A comparison of the physical properties of beech nuts and seeds revealed greater variations between batches that were harvested from similarly aged tree stands in different regions than between batches obtained from variously aged tree stands in the same region. The above implies that nut filling is more likely to be influenced by local environmental conditions than the age of the tree. Similar observations were made by KALINIEWICZ et al. (2013) in a study of pine seeds harvested from parent trees aged 124 to 180 years. The reported differences in the dimensions of pine seeds were statistically significant, but not large enough to necessitate adjustment of mesh screen parameters. Seed dimensions generally decrease with the age of parent trees. In this study, the

length of the examined nuts increased with tree age (105 to 155 years), which could be attributed to the fact that the generative phase begins relatively late in the common beech (JAWORSKI 2011), which delays the onset of old age when reproductive performance gradually declines.

Similarly to buckwheat nutlets and knotweed seeds, beech nuts are triangular, but they are much larger than the compared seeds. On average, even the lightest beech nuts are five-fold heavier than buckwheat nutlets (CAMPBELL 2003, KRAM et al. 2007) and 55-fold heavier than pale knotweed seeds (MATUSIEWICZ et al. 2014). The cross-section of common beech nuts resembles an equilateral triangle (TYLEK 2010). In this study, the differences in the width of nuts and seeds, measured on all sides and at the widest point, were statistically significant at 0.05 (Tab. 2 and 3, Fig. 5). Differences were observed in the vertical angles of nut cross-sections which were determined at 68.6, 57.5 and 53.9° in nuts and 67.3, 58.3 and 54.5° in husked seeds on average. The average width of beech nuts and seeds, measured on each of the three sides, was determined at:

- $T_1 = 7.44 \pm 0.81$ mm, $t_1 = 5.92 \pm 0.66$ mm,
- $T_2 = 8.09 \pm 0.79$ mm, $t_2 = 6.43 \pm 0.61$ mm,
- $T_3 = 8.42 \pm 0.83$ mm, $t_3 = 6.70 \pm 0.63$ mm.

Nut dimensions were highly correlated with nut mass, and similar observations were made by TYLEK (2010) and KALINIEWICZ et al. (2015). Similar correlations were noted between seed dimensions and seed mass. Nut parameters significantly influenced the characteristic traits of husked seeds.

On average, beech nuts are 1.47-times heavier than the extracted seeds, and the mass of a seed accounts for 68% of nut mass. In the common beech, the ratio of nut thickness to seed thickness is 5% higher than that reported in pumpkin seeds (JOSHI et al. 1993) and 15% higher than that noted in locust beans (OGUNJIMI et al. 2002). In the analyzed tree species, the ratio of nut width to nut length and the ratio of seed width to seed length was higher by 2% and 18% in comparison with pumpkin seeds, respectively, and by 16% and 22% in comparison with locust beans, respectively. The above results indicate that beech nuts contain more empty space than the seeds of the compared plants. In the analyzed seeds, the nut mass to seed mass ratio was approximately 17% higher than in pumpkin seeds and approximately 11% lower than in locust beans. In beech seeds, the pericarp's share of nut mass was higher than the share of the seed coat in the mass of locust beans and lower than the share of the seed coat in the mass of pumpkin seeds. The nut size to nut mass ratios and the seed size to seed mass ratios in the common beech were very similar to those reported in buckwheat nutlets (KRAM et al. 2007). The above findings indicate that beech nuts should be husked in a similar manner to buckwheat nutlets. Seeds should be divided into several size fractions, and each fraction

should be directed to a different husking device set to the corresponding parameters (JURGA 1997).

The presented correlations between the physical properties of beech nuts and seeds can be used to maximize the effectiveness of separation and husking processes, in particular during the development of separation and husking models. For a biological material, the noted results are characterized by relatively high fit to empirical data, and they can be used to improve seed treatment processes.

Conclusions

1. Despite differences in plumpness, which were conditioned by external factors and genetic traits, common beech nuts and seeds were characterized by similar cross-sectional shape. The examined cross-sections had the shape of a triangle whose sides differed significantly in length, and vertical angles were determined at 68.6, 57.5 and 53.9° in nuts and 67.3, 58.3 and 54.5° in seeds on average.

2. In both nuts and seeds, the three basic dimensions (thickness, width and length) were significantly correlated with mass, and significant correlations were also noted between the evaluated parameters. The highest value of the correlation coefficient was observed between nut mass and seed mass (approximately 0.98). The correlation coefficient was relatively high (above 0.85) in comparisons of nut and seed dimensions. Regression equations had relatively high coefficients of determination. The noted results were characterized by satisfactory fit to empirical data, and they can be used to improve the efficiency of nut cleaning and husking processes.

3. Before husking, beech nuts should be separated into size categories with the use of mesh screens with longitudinal openings. Husked seeds should be further sorted (jointly or separately for each fraction) into different mass categories, which can be achieved with the use of mesh screens with ≠6 mm and ≠7 mm openings.

References

ALTUNTAS E., DEMIRTOLA H. 2007. *Effect of moisture content on physical properties of some grain legume seeds*. New Zealand Journal of Crop and Horticultural Science, 35: 423–433. http://dx.doi.org/10.1080/01140670709510210.

BODYŁ M., SUŁKOWSKA M. 2007. *Estimation of European beech (Fagus sylvatica L.) seed differentiation in Poland in the years 1992–2004*. Sylwan, 9: 12–21.

BOLTE A., CZAKOWSKI T., KOMPA T. 2007. *The north-eastern distribution range of European beech – a review*. Forestry, 80: 413–429. http://dx.doi.org/10.1093/forestry/cpm028.

Brus D.J., Hengeveld G.M., Walvoort D.J.J., Goedhart P.W., Heidema A.H., Nabuurs G.J., Gunia K. 2012. *Statistical mapping of tree species over Europe*. European Journal of Forest Research, 131: 145–157. http://dx.doi.org/10.1007/s10342-011-0513-5.

Campbell C. 2003. *Buckwheat crop improvement*. Fagopyrum, 20: 1–6.

Dittmar C., Zech W., Elling W. 2003. *Growth variations of Common beech (Fagus sylvatica L.) under different climatic and environmental conditions in Europe – a dendroecological study*. Forest Ecology and Management, 173: 63–78. http://dx.doi.org/10.1016/S0378-1127(01)00816-7.

Institute of Dendrology in Kórnik, *Trees of the World* bookmark, common beech. www.id-pan.poznan.pl (access: 25.06.2015).

Jacobo-Valenzuela N., Moróstica-Junior M.R., Zazueta-Morales J.J., Gallegos-Infante J.A. 2011. *Physicochemical, technological properties, and health-benefits of Cucurbita moschata Duchense vs. Cehualca: A Review*. Food Research International, 44: 2587–2593. http://dx.doi.org/10.1016/j.foodres.2011.04.039.

Jaworski A. 2011. *Hodowla lasu*. Tom III. *Charakterystyka hodowlana drzew i krzewów leśnych*. PWRiL, Warszawa.

Joshi D.C., Das S.K., Mukherjee R.K. 1993. *Physical properties of pumpkin seeds*. Journal of Agricultural Engineering Research, 54(3): 219–229. http://dx.doi.org/10.1006/jaer.1993.1016.

Jurga R. 1997. *Przetwórstwo zbóż*. Część 2. WSiP, Warszawa.

Kaliniewicz Z., Grabowski A., Liszewski A., Fura S. 2011. *Analysis of correlations between selected physical attributes of Scots pine seeds*. Technical Sciences, 14(1): 13–22.

Kaliniewicz Z., Rawa T., Tylek P., Markowski P., Anders A., Fura S. 2013. *The effect of the age of Scots pine (Pinus sylvestris L.) stands on the physical properties of seeds and the operating parameters of cleaning machines*. Technical Sciences, 16(1): 63–72.

Kaliniewicz Z., Poznański A. 2013. *Variability and correlation of selected physical attributes of small-leaved lime (Tilia cordata Mill.) seeds*. Sylwan, 157(1): 39–46.

Kaliniewicz Z., Tylek P., Anders A., Markowski P., Rawa T., Andros K. 2015. *Correlations between basic physical parameters of nuts and the weight of common beech (Fagus sylvatica L.) seeds*. Technical Sciences, 18(1): 5–14.

Kram B.B., Woliński J., Wolińska J. 2007. *Comparative studies on geometric traits of nutlets with and without seed cover in Red corolla buckwheat*. Acta Agrophysica, 9(3): 657–664.

Magri D. 2008. *Patterns of post-glacial spread and the extent of glacial refugia of European beech (Fagus sylvatica)*. Journal of Biogeography, 35: 450–463. http://dx.doi.org/10.1111/j.1365-2699.2007.01803.x.

Matusiewicz M., Kubicka H., Wałejko A., Skrajna T. 2010. *Phenotypic diversity of Polygonum species in Suwalszczyzna area*. Ochrona Środowiska i Zasobów Naturalnych, 44: 58–65.

Nasiennictwo leśnych drzew i krzewów iglastych. 1995. Ed. A. Załęski. Oficyna Edytorska „Wydawnictwo Świat", Warszawa.

Ogunjimi L.A.O., Aviara N.A., Aregbesola O.A. 2002. *Some engineering properties of locust bean seed*. Journal of Food Engineering, 55: 95–99. http://dx.doi.org/10.1016/S0260-8774(02)00021-3.

Oheimb G. Von, Westphal C., Tempel H., Härdtle W. 2005. *Structural pattern of a near-natural beech forest (Fagus sylvatica) (Serrahn, North-east Germany)*. Forest Ecology and Management, 212: 253–263. http://dx.doi.org/10.1016/j.foreco.2005.03.033.

Övergaard R., Gemmel P., Karlsson M. 2007. *Effects of weather conditions on mast year frequency in beech (Fagus silvatica L.) in Sweden*. Forestry, 80(5): 555–565. http://dx.doi.org/10.1093/forestry/cpm020.

Pukacka S., Ratajczak E. 2005. *Production and scavenging of reactive oxygen species in Fagus sylvatica seeds during storage at varied temperature and humidity*. Journal of Plant Physiology, 162: 873–885. http://dx.doi.org/10.1016/j.jplph.2004.10.012.

Pukacka S., Ratajczak E. 2014. *Factors influencing the storability of Fagus sylvatica L. seeds after release from dormancy*. Plant Growth Regulation, 72: 17–27. http://dx.doi.org/10.1007/s10725-013-9832-5.

Rabiej M. 2012. *Statystyka z programem Statistica*. Helion, Gliwice.

Reyes D., Rodríguez D., Lorenzo O., Nicolás G., Cañas R., Cantón F.R., Canovas F.M., Nicolás C. 2006. *Immunolocalization of FsPK1 correlates this abscisic acid-induced protein kinase with germination arrest in Fagus sylvatica L. seeds*. Journal of Experimental Botany, 57(4): 923–929. http://dx.doi.org/10.1093/jxb/erj077.

RUSCOE W.A., ELKINTON J.S., CHOQUENOT D., ALLEN R.B. 2005. *Predation of beech seed by mice: effects of numerical and functional responses*. Journal of Animal Ecology, 74: 1005–1019. http://dx.doi.org/10.1111/j.1365-2656.2005.00998.x.

SKRZYDŁOWSKI T., PIOTROWSKI M. 2003. *Predispersal beech seed predation*. Part II. *Predation of beech seeds by vertebrates*. Sylwan, 5: 86–92.

SUSZKA B., MULLER C., BONNET-MASIMBER M. 2000. *Nasiona leśnych drzew liściastych od zbioru do siewu*. Wydawnictwo Naukowe PWN, Warszawa-Poznań.

TYLEK P. 2010. *Physical and biological aspects of the mechanical separation of seeds of the common beech (Fagus sylvatica L.)*. Zeszyty Naukowe Uniwersytetu Rolniczego im. Hugona Kołłątaja w Krakowie, 467, Rozprawy, 344, Kraków.

Technical Sciences, 2018, 21(1), 65–73

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Quarterly peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: www.uwm.edu.pl/techsci/

ANALYSIS OF CRYSTALLITE SIZE CHANGES IN A HEMATITE AND MAGNETITE FORMED ON STEEL USED IN THE POWER IDUSTRY

Monika Gwoździk

Institute of Materials Engineering
Faculty of Production Engineering and Materials Technology
Czestochowa University of Technology

Received 27 July 2017, accepted 29 November 2017, available online 6 December 2017.

K e y w o r d s: 10CrMo9-10 steel, hematite, magnetite, X-ray diffraction, crystallite sizes.

A b s t r a c t

The paper presents results of studies on the crystallite sizes of oxide layer formed during a long-term operation on 10CrMo9-10 steel at an elevated temperature (T = 545°C, t = 200,000 h). This value was determined by a method based on analysis of the diffraction line profile, according to a Scherrer formula. The oxide layer was studied on a surface and a cross-section at the outer and inner site on the pipe outlet, at the fire and counter-fire wall of the tube. X-ray studies were carried out on the surface of a tube, then the layer's surface was polished and the diffraction measurements repeated to reveal differences in the originated oxides layer.

Introduction

Steels operating at elevated temperatures (such as e.g. 10CrMo9-10, 13CrMo4-5, X10CrMoVNb9-1) are exposed to a high-temperature corrosion, which frequently results in the damage of elements operating long-term in the power industry. The originating oxide layer grows with time and increasing temperature, which then results in a loss of tube wall thickness. Such thinning of the wall occurs both on the inside and on the outside. However, on the outside the corrosion is additionally supported by aggressive compounds

Correspondence: Monika Gwoździk, Instytut Inżynierii Materiałowej, Politechnika Częstochowska, ul. Armii Krajowej 19, 42-201 Częstochowa, e-mail: gwozdzik.monika@wip.pcz.pl

existing in the flue gas (GWOŹDZIK, NITKIEWICZ 2014, GWOŹDZIK 2016a,b). Numerous papers on the oxidation of steels used in the power industry were published in recent years. The papers refer both to the short-term (up to a few hundred hours) and to the long-term oxidation (up to a few thousand or even a few hundred thousand hours) (GWOŹDZIK, NITKIEWICZ 2014, GWOŹDZIK 2016a, GWOŹDZIK 2016b, SÁNCHEZ et al. 2009). A great interest in studies on steels operated during more than 100,000 hours results from forecasting the life of thermal-mechanical equipment planned to be operated for more than 300,000 hours (ŚLIWA, GAWRON, 2010, TRZESZCZYŃSKI 2011). Both short- and long-term studies show that the oxide layer formed on steels consists of a few layers. In paper (BISCHOFF et al. 2013) performed short-term studies on two steels, HCM12A and NF616. Corrosion test were performed in steam and SCW (super critical water) at 500°C. The paper showed, that both alloys in both corrosion environments formed of Fe_3O_4 (outer layer) and a mixture of Fe_3O_4 and $FeCr_2O_4$ (inner layer). Also studies carried out by (CHEN et al. 2006) on steel NF16 and studies performed on steel X10CrMoVNb9-1 (T = 1,000 h – SÁNCHEZ et al. 2009, T = 54,144 h – GWOŹDZIK, NITKIEWICZ 2014) have shown a laminar structure of the oxide layer. At present the huge development in research equipment has taken place, especially using in materials engineering, such as: SEM, XRD, AFM, MFM, TEM (BRAMOWICZ et al. 2014, GWOŹDZIK 2016b, SZAFARSKA, IWASZKO 2012, LABISZ et al. 2016, 2017). Studies on oxidised layers are now carried out more and more often, with the application of XRD measurements. The study comprised a crystallite sizes of oxide layers formed on steels long-term operated at elevated temperatures.

Material and Experimental Methods

The material studied comprised specimens of 10CrMo9-10 steel operated at the temperature of 545°C during 200,000 h. The chemical compositions and operating parameters of steel are given in Table 1. The analysis of the chemical composition of steel was carried out using spark emission spectroscopy on a Spectrolab spectrometer.

The oxide layer was studied at the outer site (the flowing gas side) and at the inner site (the flowing steam side) on the pipe outlet, at the fire and counter-fire wall of the tube.

X-ray diffraction (XRD) measurements (studying the phase composition, crystallite sizes); the layer was subject to measurements using a Seifert 3003T/T X-ray diffractometer and the radiation originating from a tube with a cobalt anode (λ_{Co} = 0.17902 nm). X-ray studies were performed, comprising measurements in a symmetric Bragg-Brentano geometry. XRD measurements

Table 1

Chemical composition of examined steel and parameters of exploitation

Chemical composition [wt. %]							
Acc.	C	Si	Mn	P	S	Cr	Mo
Analysis	0.11	0.34	0.55	0.002	0.009	2.23	0.98
EN	0.08–0.14	max. 0.50	0.40–0.80	max. 0.020	max. 0.010	2.00–2.50	0.90–1.10
Parameters of exploitation							
Temperature [°C]				time [h]			
545				200,000			

in the first stage (surface measurements and after the first polishing) were carried out in the range of 20÷120° angle for phase identification which were presented in paper (GWOŹDZIK 2016b), then the measurements were narrowed to an angle range of 35÷45° with an angular step of 0.1° and exposure time 4 s. To interpret the results the diffractograms were described by a Pseudo Voigt curve using the Analyze software. A Pseudo Voigt function was used in order to accurately determine the location of the major diffraction reflections. A computer software and the PDF4+2009 crystallographic database were used for the phase identification.

Based on the width and the position of the main coat and substrate reflections, the size of the crystallites was determined using the Scherrer formula (1) (CULLITY 1964, GWOŹDZIK 2016c):

$$D_{hkl} = \frac{\mathrm{k} \cdot \lambda}{\beta \cdot \cos\theta} \tag{1}$$

where:
D_{hkl} – crystallite size in the direction normal to (hkl) [nm],
k – constant (~1);
λ – radiation wavelength [nm],
β – reflection width depending on the crystallite size [rad],
θ – Bragg angle [rad].

X-ray studies were carried out on the surface, and then the layer surface was polished down and the diffraction measurements were performed again to determine individual oxide layers.

The size of the D_{hkl} crystallites size was determined for the reflections originating from the planes (104) for Fe_2O_3 and (311) for Fe_3O_4, which are occurring at angles of 38.7464° and 40.8998°, respectively (according to the catalog card ICDD PDF 01-079-0007 and ICDD PDF 01-089-0951). The catalog standards for Fe_2O_3 and Fe_3O_4 have been shown on the Figure 1.

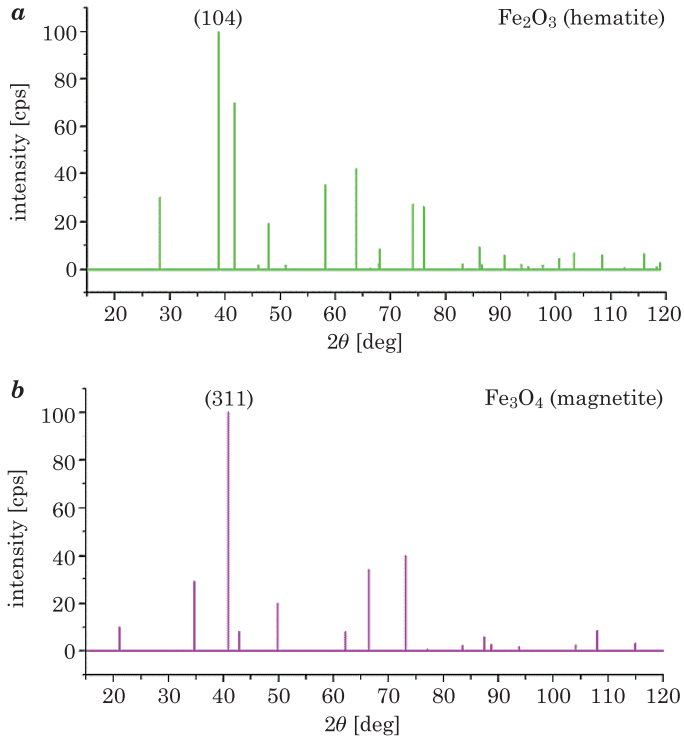

Fig. 1. Standards corresponding to individual oxides, in which the main reflections originating from planes are marked

X-ray measurements were performed at different depths of the oxide layer.

After removal of the sediment layer (400 μm, 10 μm, 6 μm and 4 μm, for outer site-fire wall, outer site-opposite fire wall, inner site-fire wall, inner site-opposite fire wall, respectively), X-ray measurements were carried out, then for each case, the oxide layer was removed (20 μm) cyclically, each time making XRD measurements:

– from exhaust side – fire side, the oxide layer was removed every 20 μm in 21 cycles,

– from exhaust side – counter-fire side, the oxide layer was removed every 20 μm in 18 cycles,

– from steam side – fire side, the oxide layer was removed every 20 μm in 21 cycles,

– from steam side – counter-fire side, the oxide layer was removed every 20 μm in 20 cycles.

The exemplary of examine of XRD measurements has been shown in Figure 2.

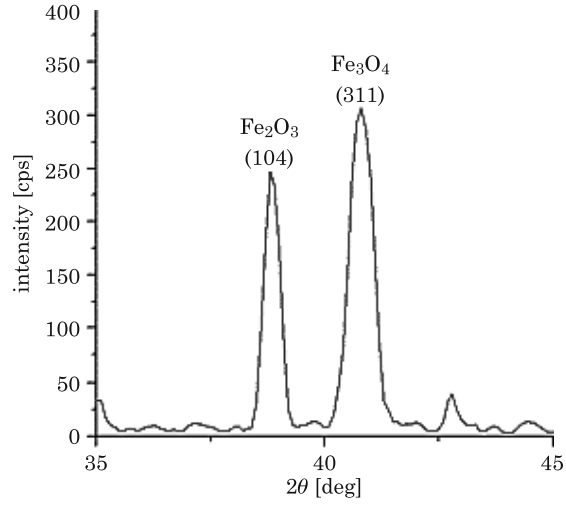

Fig. 2. X-ray diffraction patterns from the oxides layer obtained by means of XRD technique

Results of examinations

The obtained results have shown that in the case of the fire side of the flue gas side after removing the correct the oxide layer by 20 μm, the size of the D_{hkl} crystallites for hematite was the highest and it equals 49 nm (Fig. 3). Further successive removal of the hematite layer showed a decrease of crystallite size. For magnetite, the gradual increase was observed, the next delicate decrease and the next increase of crystallite sizes has been observed together with the

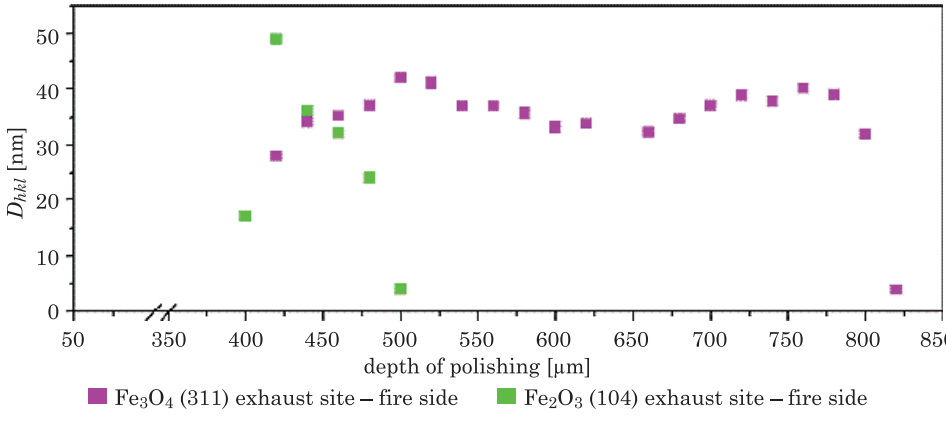

Fig. 3. Determination of crystallite size D_{hkl} for main peaks Fe_3O_4 and Fe_2O_3, exhaust side – fire side

depth of polishing. The highest value of this parameter was obtained at a depth 500 μm of oxide layer, where D_{hkl} was 42 nm.

In the case of the opposite-fire side of the flue gas side (Fig. 4), the D_{hkl} for hematite have been in the range of about 48 nm at a depth of 30 μm. The sharp of decrease this parameter was observed at a depth of 50 μm. In case of magnetite, the D_{hkl} is oscillated around 40 nm in a depth of oxide layer from 50 to 170 μm. It has been shown decrease of crystallite size at a depth from 210 μm to 230 μm and increase of D_{hkl} in depth 250 μm.

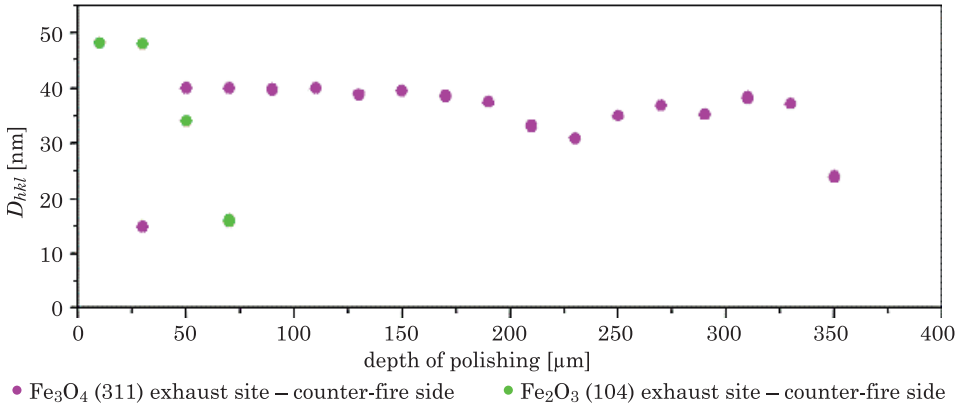

Fig. 4. Determination of crystallite size D_{hkl} for main peaks Fe_3O_4 and Fe_2O_3, exhaust side – counter-fire side

The size of crystallites determined based on the Scherrer formula on the steam side shows smaller dimensions than from flue gas side. The parameter of D_{hkl} of hematite (for steam side) the largest values equal 46 nm and 44 nm for the fire side and the opposite fire side respectively (Fig. 5 and 6). In both cases, this parameter decreases dramatically in the depth of the layer is greater than 86 μm. On the fire side, the polishing (up to 186 μm) has shown a delicate growth of crystallites of magnetite, the next series of polishing (from 186 μm to 400 μm) has shown as small decrease of crystallite size. For the fire side, crystallite sizes of magnetite has been increasing up to a depth of 84 μm. The next polishing (from 84 μm to 164 μm) have shown that the crystallite size is maintained at one level. The polishing depth above 164 μm has shown a decrease in D_{hkl} which oscillates around the same value up to a depth of polishing 380 μm.

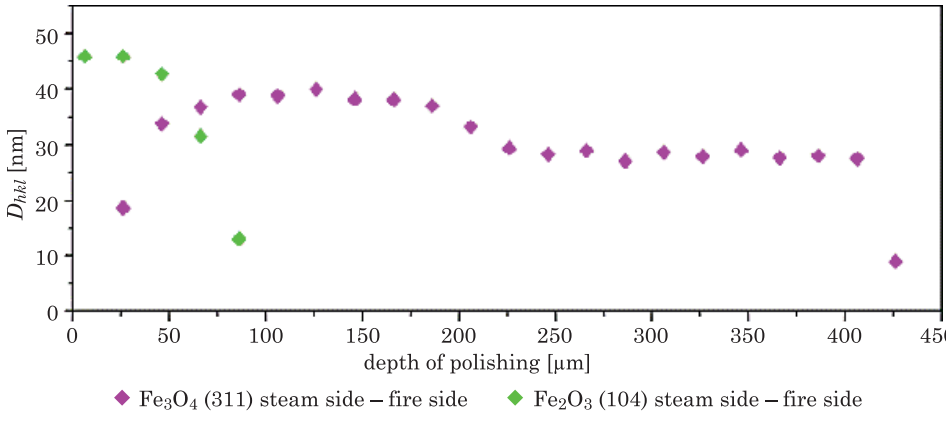

Fig. 5. Determination of crystallite size D_{hkl} for main peaks Fe_3O_4 and Fe_2O_3, steam side – fire side

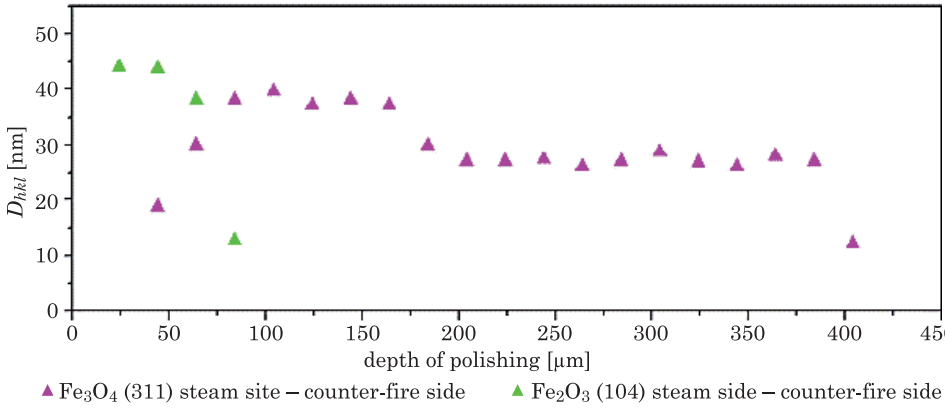

Fig. 6. Determination of crystallite size D_{hkl} for main peaks Fe_3O_4 and Fe_2O_3, steam side – counter-fire side

Summary

The paper presents results of studies on oxides formed on 10CrMo9-10 steel operated at the temperature of 545°C during 200,000 h. The oxide layer formed on the flue gas side (the outer side) and steam side (the inner side) both on the fire and counter-fire side were analyzed.

The investigated of oxide layer formed on this steel have shown more degraded of hematite and magnetite in the case of larger crystallites. The results obtained are well correlated with the results obtained with the use of an

optical microscope, as have shown in paper (GWOŹDZIK 2016b). The oxide layer thickness together with deposits on the fire side was 435 μm and 820 μm on the flowing medium and flue gas side, respectively. For the opposite fire side on the inside the oxides layer thickness was 405 μm while on the outside 360 μm. Examinations carried out have shown that the formed layer is the most degraded on the fire side from flue gas side. The size of crystallites on this side have been the largest dimensions both for hematite and magnetite. Instead, comparing the size of hematite and magnetite crystallites it is possible to state that in every cases D_{hkl} is larger for hematite. The size of crystallites in oxides originating during long-term operation depends on oxides morphology, where it has been presented in previous paper the author. In paper (GWOŹDZIK 2016c), the author showed that the size of crystallites determined based on the Scherrer formula on the fire side shows much larger dimensions for oxides. Instead, comparing the size of hematite and magnetite crystallites it is possible to state that D_{hkl} is larger for hematite.

References

BISCHOFF J., MOTTA A.T., EICHFELD C., COMSTOCK R.J., CAO G., ALLEN T.R. 2013. *Corrosion of ferritic-martensitic steels in steam and supercritical water.* Journal of Nuclear Materials, 441: 604–611.

BRAMOWICZ M., KULESZA S., MROZEK G. 2014. *Changes in magnetic domain structure of maraging steel studied by magnetic force microscopy.* Technical Sciences, 17(4): 371–379.

CHEN Y., SRIDHARAN K., ALLEN T.R., UKAI S. 2006. *Mictrostructural examination of oxide layers formed on an oxide dispersion strengthened ferritic steel exposed to supercritical water.* Journal of Nuclear Materials, 359: 50–58.

CULLITY B.D. 1964. *Podstawy dyfrakcji promieni rentgenowskich.* PWN, Warszawa.

EN 10028-2 *Flat products made of steels for pressure purposes*. Part 2. *Non-alloy and alloy steels with specified elevated temperature properties*.

GWOŹDZIK M., NITKIEWICZ Z. 2014. *Studies on the adhesion of oxide layer formed on X10CrMoVNb9-1 steel.* Archives of Civil and Mechanical Engineering, 14: 335–341.

GWOŹDZIK M. 2016a. *Structure studies of porous oxide layers formed on 13CrMo4-5 steels long-term operated in the power industry.* Technical Sciences, 19(3): 257–266.

GWOŹDZIK M. 2016b. *The defects of oxide layers formed on 10CrMo9-10 steel operated for 200,000 hours at an elevated temperature.* Archives of Metallurgy and Materials, 2B(61): 987–992.

GWOŹDZIK M. 2016c. *Analysis of Crystallite Size Changes in an Oxide Layer Formed on Steel Used in the Power Industry.* Acta Physica Polonica A, 4(130): 935–938.

LABISZ K., KRUPIŃSKI M., PAWLYTA M., MATUS K., KREMZER M., DOPIERALA K. 2016. *High Resolution TEM Investigations and TDA Analysis of Zinc Alloy with Strontium Addition.* Acta Physica Polonica A, 130(4): 823–826.

LABISZ K., TAŃSKI T., KREMZER M., JANICKI D. 2017. *Effect of laser alloying on heat-treated light alloys.* International Journal of Materials Research, 108(2): 126–132.

SZAFARSKA M., IWASZKO J. 2012. *Laser remelting treatment of plasma-sprayed Cr2*O3 oxide coatings. Archives of Metallurgy and Materials, 57(1): 215–221.

SÁNCHEZ L., HIERRO M.P., PÉREZ F.J. 2009. *Effect of chromium content on the oxidation behaviour of ferritic steels for applications in steam atmospheres at high temperatures.* Oxidation of Metals, 71: 173–186.

Śliwa A., Gawron P. 2010. *Wpływ warunków chemicznych pracy bloków energetycznych na możliwość przedłużania ich eksploatacji powyżej 300000 godzin*. XII Sympozjum Informacyjno-Szkoleniowe: „Diagnostyka i remonty urządzeń cieplno-mechanicznych elektrowni. Modernizacje urządzeń energetycznych w celu przedłużenia ich eksploatacji powyżej 300000 godzin", Wisła, p. 23–28.

Trzeszczyński J. 2011. *System diagnostyczny zapewniający bezpieczną pracę bloków 200MW eksploatowanych powyżej 300000 godzin*. XIII Sympozjum Informacyjno-Szkoleniowe: „Diagnostyka i remonty urządzeń cieplno-mechanicznych elektrowni. Zarządzanie majątkiem produkcyjnym grupy elektrowni", Katowice, p. 87–90.

Guide for Autors

Introduction

Technical Sciences is a peer-reviewed research Journal published in English by the Publishing House of the University of Warmia and Mazury in Olsztyn (Poland). Journal is published continually since 1998. Until 2010 Journal was published as a yearbook, in 2011 and 2012 it was published semiyearly. From 2013, the Journal is published quarterly in the spring, summer, fall, and winter.

The Journal covers basic and applied researches in the field of engineering and the physical sciences that represent advances in understanding or modeling of the performance of technical and/or biological systems. The Journal covers most branches of engineering science including biosystems engineering, civil engineering, environmental engineering, food engineering, geodesy and cartography, information technology, mechanical engineering, materials science, production engineering etc.

Papers may report the results of experiments, theoretical analyses, design of machines and mechanization systems, processes or processing methods, new materials, new measurements methods or new ideas in information technology.

The submitted manuscripts should have clear science content in methodology, results and discussion. Appropriate scientific and statistically sound experimental designs must be included in methodology and statistics must be employed in analyzing data to discuss the impact of test variables. Moreover there should be clear evidence provided on how the given results advance the area of engineering science. Mere confirmation of existing published data is not acceptable. Manuscripts should present results of completed works.

There are three types of papers: a) research papers (full length articles); b) short communications; c) review papers.

The Journal is published in the printed and electronic version. The electronic version is published on the website ahead of printed version of Technical Sciences.

Technical Sciences does not charge submission or page fees.

Types of paper

The following articles are accepted for publication:

Reviews

Reviews should present a focused aspect on a topic of current interest in the area of biosystems engineering, civil engineering, environmental engineering, food engineering, geodesy and cartography, information technology, mechanical engineering, materials science, production engineering etc. They should include all major findings and bring together reports from a number of sources. These critical reviews should draw out comparisons and conflicts between work, and provide an overview of the 'state of the art'. They should give objective assessments of the topic by citing relevant published work, and not merely present the opinions of individual authors or summarize only work carried out by the authors or by those with whom the authors agree. Undue speculations should also be avoided. Reviews generally should not exceed 6,000 words.

Research Papers

Research Papers are reports of complete, scientifically sound, original research which contributes new knowledge to its field. Papers should not exceed 5,000 words, including figures and tables.

Short Communications

Short Communications are research papers constituting a concise description of a limited investigation. They should be completely documented, both by reference list, and description of the experimental procedures. Short Communications should not occupy more than 2,000 words, including figures and tables.

Letters to the Editor

Letters to the Editor should concern with issues raised by articles recently published in scientific journals or by recent developments in the engineering area.

Contact details for submission

The paper should be sent to the Editorial Office, as a Microsoft Word file, by e-mail: techsci@uwm.edu.pl

Referees

Author/authors should suggest, the names, addresses and e-mail addresses of at least three potential referees. The editor retains the sole right to decide whether or not the suggested reviewers are used.

Submission declaration

After final acceptance of the manuscript, the corresponding author should send to the Editorial Office the author's declaration. Submission of an article implies that the work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint), that it is not under consideration for publication elsewhere, that publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language.

To prevent cases of ghostwriting and guest authorship, the author/authors of manuscripts is/are obliged to: (i) disclose the input of each author to the text (specifying their affiliations and contributions, i.e. who is the author of the concept, assumptions, methods, protocol, etc. used during the preparation of the text); (ii) disclose information about the funding sources for the article, the contribution of research institutions, associations and other entities.

Language

Authors should prepare the full manuscript i.e. title, abstract and the main text in English (American or British usage is accepted). Polish version of the manuscript is not required.

The file type

Text should be prepared in a word processor and saved in doc or docx file (MS Office).

Article structure

Suggested structure of the manuscript is as follows:

Title
Authors and affiliations
Corresponding author
Abstract
Keywords
Introduction
Material and Methods
Results and Discussion
Conclusions

Acknowledgements (*optional*)
References
Tables
Figures

Subdivision – numbered sections

Text should be organized into clearly defined and numbered sections and subsections (optionally). Sections and subsections should be numbered as 1. 2. 3. then 1.1 1.2 1.3 (then 1.1.1, 1.1.2, ...). The abstract should not be included in numbering section. A brief heading may be given to any subsection. Each heading should appear on its own separate line. A single line should separate paragraphs. Indentation should be used in each paragraph.

Font guidelines are as follows:

- Title: 14 pt. Times New Roman, bold, centered, with caps
- Author names and affiliations:12 pt. Times New Roman, bold, centered, italic, two blank line above
- Abstract: 10 pt. Times New Roman, full justified, one and a half space. Abstract should begin with the word Abstract immediately following the title block with one blank line in between. The word Abstract: 10 pt. Times New Roman, centered, indentation should be used
- Section Headings: Not numbered, 12 pt. Times New Roman, bold, centered; one blank line above
- Section Sub-headings: Numbered, 12 pt. Times New Roman, bold, italic, centered; one blank line above
- Regular text: 12 pt. Times New Roman, one and a half space,full justified, indentation should be used in each paragraph

Title page information

The following information should be placed at the first page:

Title

Concise and informative. If possible, authors should not use abbreviations and formulae.

Authors and affiliations

Author/authors' names should be presented below the title. The authors' affiliation addresses (department or college; university or company; city, state and zip code, country) should be placed below the names. Authors with the same affiliation must be grouped together on the same line with affiliation information following in a single block. Authors should indicate all affiliations with a lower-case superscript letter immediately after the author's name and in front of the appropriate address.

Corresponding author

It should be clearly indicated who will handle correspondence at all stages of refereeing and publication, also post-publication process. The e-mail address should be provided (footer, first page). Contact details must be kept up to date by the corresponding author.

Abstract

The abstract should have up to 100-150 words in length. A concise abstract is required. The abstract should state briefly the aim of the research, the principal results and major conclusions. Abstract must be able to stand alone. Only abbreviations firmly

established in the field may be eligible. Non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself.

Keywords

Immediately after the abstract, author/authors should provide a maximum of 6 keywords avoiding general, plural terms and multiple concepts (avoid, for example, 'and', 'of'). Author/authors should be sparing with abbreviations: only abbreviations firmly established in the field may be eligible.

Abbreviations

Author/authors should define abbreviations that are not standard in this field. Abbreviations must be defined at their first mention there. Author/authors should ensure consistency of abbreviations throughout the article.

Units

All units used in the paper should be consistent with the SI system of measurement. If other units are mentioned, author/authors should give their equivalent in SI.

Introduction

Literature sources should be appropriately selected and cited. A literature review should discuss published information in a particular subject area. Introduction should identify, describe and analyze related research that has already been done and summarize the state of art in the topic area. Author/authors should state clearly the objectives of the work and provide an adequate background.

Material and Methods

Author/authors should provide sufficient details to allow the work to be reproduced by other researchers. Methods already published should be indicated by a reference. A theory should extend, not repeat, the background to the article already dealt within the Introduction and lay the foundation for further work. Calculations should represent a practical development from a theoretical basis.

Results and Discussion

Results should be clear and concise. Discussion should explore the significance of the results of the work, not repeat them. A combined Results and Discussion section is often appropriate.

Conclusions

The main conclusions of the study may be presented in a Conclusions section, which may stand alone or form a subsection of a Results and Discussion section.

Acknowledgements

Author/authors should include acknowledgements in a separate section at the end of the manuscript before the references. Author/authors should not include them on the title page, as a footnote to the title or otherwise. Individuals who provided help during the research study should be listed in this section.

Artwork
General points

– Make sure you use uniform lettering and sizing of your original artwork
– Embed the used fonts if the application provides that option

– Aim to use the following fonts in your illustrations: Arial, Courier, Times New Roman, Symbol
– Number equations, tables and figures according to their sequence in the text
– Size the illustrations close to the desired dimensions of the printed version

Formats

If your electronic artwork is created in a Microsoft Office application (Word, PowerPoint, Excel) then please supply 'as is' in the native document format

Regardless of the application used other than Microsoft Office, when your electronic artwork is finalized, please 'Save as' or convert the images to one of the following formats (note the resolution requirements given below):

EPS (or PDF): Vector drawings, embed all used fonts

JPEG: Color or grayscale photographs (halftones), keep to a minimum of 300 dpi

JPEG: Bitmapped (pure black & white pixels) line drawings, keep to a minimum of 1000 dpi or combinations bitmapped line/half-tone (color or grayscale), keep to a minimum of 500 dpi

Please **do not**:
– Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); these typically have a low number of pixels and limited set of colors
– Supply files that are too low in resolution
– Submit graphics that are disproportionately large for the content

Color artwork

Author/authors should make sure that artwork files are in an acceptable format (JPEG, EPS PDF, or MS Office files) and with the correct resolution. If, together with manuscript, author/authors submit color figures then Technical Sciences will ensure that these figures will appear in color on the web as well as in the printed version at no additional charge.

Tables, figures, and equations

Tables, figures, and equations/formulae should be identified and numbered consecutively in accordance with their appearance in the text.

Equations/mathematical and physical formulae should be presented in the main text, while tables and figures should be presented at the end of file (after References section). Mathematical and physical formulae should be presented in the MS Word formula editor.

All types of figures can be black/white or color. Author/authors should ensure that each figure is numbered and has a caption. A caption should be placed below the figure. Figure must be able to stand alone (explanation of all symbols and abbreviations used in figure is required). Units must be always included. It is noted that figure and table numbering should be independent.

Tables should be numbered consecutively in accordance with their appearance in the text. Table caption should be placed above the table. Footnotes to tables should be placed below the table body and indicated with superscript lowercase letters. Vertical rules should be avoided. Author/authors should ensure that the data presented in tables do not duplicate results described in figures, diagrams, schemes, etc. Table must be able to stand alone (explanation of all symbols and abbreviations used in table is required). Units must be always included. As above, figure and table numbering should be independent.

References

References: All publications cited in the text should be presented in a list of references following the text of the manuscript. The manuscript should be carefully checked to ensure that the spelling of authors' names and dates of publications are exactly the same in the text as in the reference list. Authors should ensure that each reference cited in the text is also present in the reference list (and vice versa).

Citations may be made directly (or parenthetically). All citations in the text should refer to:

1. Single author

The author's name (without initials, with caps, unless there is ambiguity) and the year of publication should appear in the text

2. Two authors

Both authors' names (without initials, with caps) and the year of publication should appear in the text

3. Three or more authors

First author's name followed by et al. and the year of publication should appear in the text

Groups of references should be listed first alphabetically, then chronologically.

Examples:

"... have been reported recently (ALLAN, 1996a, 1996b, 1999; ALLAN and JONES, 1995). KRAMER et al. (2000) have recently shown..."

The list of references should be arranged alphabetically by authors' names, then further sorted chronologically if necessary. More than once reference from the same author(s) in the same year must be identified by the letters "a", "b", "c" etc., placed after the year of publication.

References should be given in the following form:

KUMBHAR B.K., AGARVAL R.S., DAS K. 1981. *Thermal properties of fresh and frozen fish*. International Journal of Refrigeration, 4(3), 143–146.

MACHADO M.F., OLIVEIRA F.A.R., GEKAS V. 1997. *Modelling water uptake and soluble solids losses by puffed breakfast cereal immersed in water or milk*. In Proceedings of the Seventh International Congress on Engineering and Food, Brighton, UK.

NETER J., KUTNER M.H., NACHTSCHEIM C.J., WASSERMAN W. 1966. *Applied linear statistical models* (4th ed., pp. 1289–1293). Irwin, Chicago.

THOMSON F.M. 1984. *Storage of particulate solids*. In M. E. Fayed, L. Otten (Eds.), Handbook of Powder Science and Technology (pp. 365–463). Van Nostrand Reinhold, New York.

Citation of a reference as 'in press' implies that the item has been accepted for publication.

Note that the full names of Journals should appear in reference list.

Submission checklist

The following list will be useful during the final checking of an article prior to the submission. Before sending the manuscript to the Journal for review, author/authors should ensure that the following items are present:

– Text is prepared with a word processor and saved in DOC or DOCX file (MS Office). One author has been designated as the corresponding author with contact details: e-mail address

– Manuscript has been 'spell-checked' and 'grammar-checked'

– References are in the correct format for this Journal

– All references mentioned in the Reference list are cited in the text, and vice versa

– Author/authors does/do not supply files that are too low in resolution

– Author/authors does/do not submit graphics that are disproportionately large for the content