

# Web Server Routine for Materials Properties Correlation Study. Theory and Applications for a set of alloys

L. Jäntschi

*Faculty of Science and Engineering of Materials, Technical University of Cluj-Napoca, Romania,  
lori@lori.east.utcluj.ro*

## Abstract

The present paper is focused on modeling of statistical data processing with applications in field of material science and engineering. A new method of data processing is presented and applied on a set of 10 Ni–Mn–Ga ferromagnetic ordered shape memory alloys that are known to exhibit phonon softening and soft mode condensation into a premartensitic phase prior to the martensitic transformation itself. The method allows to identify the correlations between data sets and to exploit them later in statistical study of alloys. An algorithm for computing data was implemented in preprocessed hypertext language (PHP) and a hypertext markup language interface for them was also realized and put onto comp.east.utcluj.ro educational web server and it is accessible via http protocol at the address <http://comp.east.utcluj.ro/~lori/research/regression/linear/v1.4/>. The program running for the set of alloys allow to identify groups of alloys properties and give qualitative measure of correlations between properties. Surfaces of property dependencies are also fitted.

*Keywords:* Shape memory alloys, Modeling, Analytical methods, Automat processing of data.

## 1. INTRODUCTION

Many statistical procedures for processing data are now available [1]. Most of them offer a voluble set of possibilities and variants, but which one to consider them? That is not a easy question and the frequent answer is: that is choice of analyst [2,3].

Data mining technology offer in this area of knowledge some answers, but not a complete answer [4]. By other hand, to interpret experiment results, data need to be well processed [5].

Structure investigations are frequently combined with statistical processing [6]. In most of cases, best results are obtained with specific procedures in contrast to general numeric algorithms [7,8]. Modeling of structure is benefit to property predictions [9,10]. Nonstandard statistical evaluation procedures then are helpful [11].

The presented model make data preprocessing to a set of 10 Ni–Mn–Ga ferromagnetic ordered shape memory alloys that are known to exhibit phonon softening and soft mode condensation into a premartensitic phase prior to the martensitic transformation itself and is a extension added to the model presented in book [12].

## 2. METHOD

The logic scheme of data preprocessing is presented in figure 1.

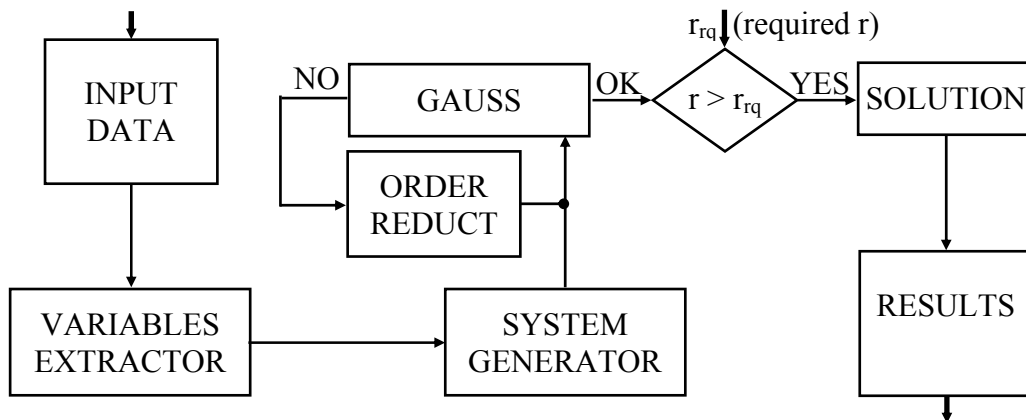


Fig. 1. Data automat processing algorithm

The INPUT module read a text format data, process input data, split it into rows and columns and computes average means.

If name  $n\_rows$  it assigned to number of rows,  $n\_cols$  to number of columns,  $data$  to array of data, the output of module INPUT is computed by formulas:

$$M[i,j] = \frac{\sum_{k=1}^{n\_rows} data[k][i] \cdot data[k][j]}{n\_rows}; M[0,j] = \frac{\sum_{k=1}^{n\_rows} data[k][i]}{n\_rows}, 1 \leq i, j \leq n\_cols \quad (1)$$

Linear regression and PLS (partial least squares) are most used methods in statistical processing of data. Presented method uses them.

The output of INPUT module is used as input in GAUSS and RESULTS modules.

GAUSS module solves a linear system of equations in form:

$$\sum_{j=1}^{n\_cols} M[i,j] \cdot x_j = 1, 1 \leq i \leq n\_cols \quad (2)$$

If answer of algorithm solving is *undetermined system* and null variable is  $x_{n\_cols}$  then GAUSS module solve determined system of  $n\_cols$  order given by equation (3):

$$\sum_{j=1}^{n\_cols-1} M[i,j] \cdot x_j = M[0,j], 1 \leq i < n\_cols \quad (3)$$

If answer of algorithm solving is *undetermined system* and null variable is different form  $x_{n\_cols}$  then GAUSS module pass extended system matrix to REDUCT ORDER module.

If is input in REDUCT ORDER module then is an *undetermined system* and this it extract null row and column corresponding to the null variable (figure 2) and the resulting matrix of  $(n\_cols-1) \times n\_cols$  dimension is passed again to GAUSS module.

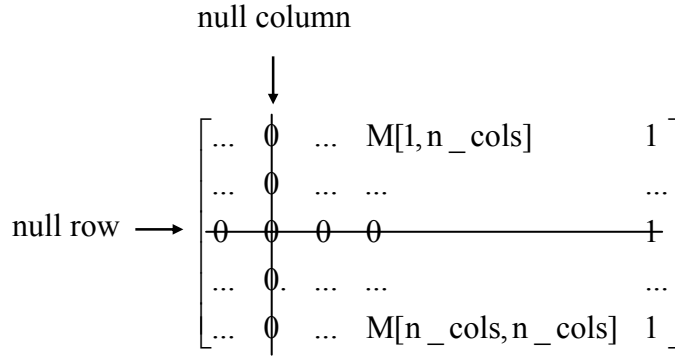


Fig. 2. Processing data in REDUCT ORDER module

When system is solved a unique solution is found. Then, system extended matrix contain at column  $n\_cols$  the coefficients of regression equation:

$$a_1 \cdot x_1 + \dots + a_i \cdot x_i + \dots + a_{n\_cols} \cdot x_{n\_cols} + a_{n\_cols+1} = 0 \quad (4)$$

where the coefficients  $a_{n\_cols+1}$  and  $a_{n\_cols+1}$  are resulted regression coefficients. Note that equation (4) is in implicit form; to obtain an explicit form is necessary to extract dependent variable from (4). The last coefficient is assigned to -1:

$$a_{n\_cols+1} = -1 \quad (5)$$

At the end of module SOLUTION it result an implicit linear regression equation between given variables through his values in columns (equation 4). Equation 4 can be exploited to obtain explicit linear regression equations for each variable which has no null coefficient  $a_i$ :

$$\hat{x}_i = \left( \frac{a_1}{-a_i} \right) \cdot x_1 + \dots + \left( \frac{a_{i-1}}{-a_i} \right) \cdot x_{i-1} + \left( \frac{a_{i+1}}{-a_i} \right) \cdot x_{i+1} + \dots + \left( \frac{a_{n\_cols}}{-a_i} \right) \cdot x_{n\_cols} + \left( \frac{a_{n\_cols+1}}{-a_i} \right) \quad (6)$$

Sum of residues  $S_i$  can be now evaluated:

$$S_i = \left( \frac{a_{n\_cols+1}}{a_i} + \sum_{j=1}^{n\_cols} \frac{a_j}{a_i} \cdot x_j \right)^2 \quad (7)$$

To compare one equation to another, an order value is required. Let to explicit this. If  $x_1$  values (data[1] from input) are percents expressed in values from 0 to 100 and  $x_2$  are preartensitic temperatures transformation expressed in K with values from 100 to 600, then also sum of residues are expressed square of same measurement units. To make independence of measurement unit and measure order, values  $S_i$  are divided with own sum of squares of variable measurements ( $M[i,i]$  from INPUT module, equation 1). Final equation, with substitution  $x_i = data[k,i]$ ,  $1 \leq k \leq n\_rows$  and summing is:

$$Q_i = \sum_{k=1}^{n\_rows} \left( \frac{a_{n\_cols+1}}{a_i} + \sum_{j=1}^{n\_cols} \frac{a_j}{a_i} \cdot data[k,i] \right)^2 / M[i,i] \quad (9)$$

and express relative residues of variable  $x_i$  when variable  $x_i$  is assumed to be dependent of independent variables  $x_1, \dots, x_{i-1}, x_{i+1}, x_{\text{cols}}$ . Note that the dependence and independence statistical concept is hard to prove in practical situations, but will see later, can be decelerated. For a good correlation,  $Q_i$  should be smallest possible value.

Another quantitative measure for a good correlation is correlation coefficient between measured data  $x_i$  and estimated values  $\hat{x}_i$  from equation (6). Assuming that  $M$  is mean operator, for  $r$  is given by:

$$r(x_i, \hat{x}_i) = \frac{M(x_i \cdot \hat{x}_i) - M(x_i) \cdot M(\hat{x}_i)}{(M(x_i \cdot x_i) - M(x_i) \cdot M(x_i))^{1/2} \cdot (M(\hat{x}_i \cdot \hat{x}_i) - M(\hat{x}_i) \cdot M(\hat{x}_i))^{1/2}} \quad (10)$$

The absolute value of  $r$  must be high for a good correlation. More additionally tests are also available in other programs such as Microsoft Excell or Statsoft Statistica.

### 3. ALGORITHM AND IMPLEMENTATION

The implementation of algorithm is relative simple, if are used a flexible language processing. In terms of programming, portability of resulted program can be a problem. As example, if we are chose to implement the algorithm in Visual Basic, the execution of the program is restricted to Windows machines. If Perl is our choice, a Unix-based machine is necessary to run program. Even if we choose to implement the program in C language, we will have serious difficulties to compile the programs on machines running with different operating systems.

Other questions require an answer: We want a server based application or client based application? We want a server side application or a client side application?

As example, a client side application can have disadvantage of execution on client, and dependence of processing speed by power of client machine. If we prefer this variant, a java script or visual basic script is our programming language.

Most benefit to portability and execution speed seems to be a PHP (post processed hypertext) variant of implementation. A PHP script can be put on any server or client with PHP processor and executed from them trough http server (Apache, Squid, ...) and client (Internet Explorer, Netscape).

Another advantage of using PHP is the possibility to link our algorithm with a materials database (d-Base, Interbase, MySQL, PostGRES format) and input data can be then loaded from them.

As conclusion, a PHP implementation is our choice.

A graphical interface was built in html with a TEXTAREA for input data and an INPUT SUBMIT button for submitting data to the server. The server is a Free BSD Unix based server with an Apache web server running on. The server is hosted in educational network of Technical University of Cluj-Napoca with address 193.226.7.211 and name comp.east.utcluj.ro.

With PHP technology, was build a routine for pseudo domain names, that redirect client to different pages, depending on his input of domain name in client http browser.

The program build have 21 subroutines and a main program, specified below:

- function af\_ec(\$n,\$coef,&\$t) // display an equation;
- function af\_mt(\$titlu,&\$stabel,\$n\_r,\$n\_c) //display any matrix with a title;
- function af\_rez(\$n\_r,\$n\_c,&\$d,&\$m,&\$c,&\$t,\$n\_o,\$pr) //list a table with best equations founded;

- function ch\_ln(\$l1,\$l2,&\$cc,\$r) // Gauss linear algebra method, change two lines in system extended matrix;
- function cnk(\$k,\$n\_r,\$n\_c,&\$data,&\$tab,\$pr,&\$dep,&\$inv) //make recursive all possible combination c(n,k);
- function data\_copy(\$n\_r,\$n\_c,&\$d,&\$t,&\$d\_t,&\$n\_t) //extract a subset of data from entire set;
- function do\_means(&\$data,&\$mean,\$n\_rows,\$n\_cols) //make all (xi, xi\*xj) means;
- function ec\_by(\$n,&\$coef,\$by,&\$c\_by) //calculate coefficients for explicit equation from implicit equation coefficients;
- function ec\_val(\$n,&\$valori,&\$coef) //compute value of implicit equation in given point
- function estimare(\$n\_r,\$n\_c,&\$d,&\$c,\$x,&\$x\_est) //compute value for explicit equation in given point for given dependent variable;
- function im\_ln(\$nr,\$rw,&\$cc,\$r) //Gauss linear algebra method, make a unitary element into system extended matrix;
- function mx\_rw(\$cl,&\$cc,\$r) //Gauss linear algebra method, find the best line for zeroes in system extended matrix;
- function n\_to\_s(\$nr) //format and display a real number;
- function r\_stat(\$n,\$k,&\$d1,&\$d2) //compute correlation coefficient r;
- function rd\_gs(\$cc,\$r,&\$cf) //Gauss linear algebra iterative algorithm;
- function reg\_lin\_1(\$n\_r,\$n\_c,&\$d,&\$t,\$pr) //make linear regression if possible; return answer;
- function res(&\$t,\$n) //reset counter for recursive c(n,k);
- function sum\_r(\$n\_rows,\$n\_cols,&\$data,&\$coef,\$cor) //calculate sum of residues;
- function ze\_pd(&\$cc,\$r) //Gauss linear algebra method, make supdiag. zeroes in system matrix;
- function ze\_sd(\$e,&\$cc,\$r) //Gauss linear algebra method, make subdiag. zeroes in system matrix;
- main program //input data and requested minimal correlation coefficient and display founded equations;

#### 4. RESULTS AND DISCUSSION

A set of Ni–Mn–Ga ferromagnetic ordered shape memory alloys are used for investigation [13]. The properties are described in table 2.

Table 2. Processed Data

Column	Property	Measurement unit
1	Alloy State (Poly- or Single-crystalline alloy)	1, -1 (PC, SC respectively)
2	e/a	Electron/atom ratio
3	Concentration of Ni	%
4	Concentration of Mn	%
5	Concentration of Ga	%
*6	T <sub>1</sub> (rows 1-7), T <sub>M</sub> ' (rows 8-10)	K
7	T <sub>M</sub> , premartensitic temperature transformation	K

\*temperatures: T<sub>1</sub>= martensitic transformation; T<sub>M</sub>'=intermartensitic transformation in Group III alloys.

Table 3. Input data values (output by PHP program)

0	1	2	3	4	5	6	7		0	1	2	3	4	5	6	7
1	1	7.35	49.6	21.9	28.5	4.2	178		6	1	7.56	47.7	30.5	21.8	227	240
2	1	7.36	47.6	25.7	26.7	4.2	152		7	1	7.57	51.1	24.9	24	197	248
3	-1	7.45	49.7	24.3	26	183	218		8	-1	7.78	53.1	26.6	20.3	417	379
4	1	7.5	50.9	23.4	25.7	113	224		9	-1	7.83	51.2	31.1	17.7	443	415
5	-1	7.51	49.2	26.6	24.2	184	240		10	-1	7.91	59	19.4	21.6	633	517

Program computes and output the regression equations. With an  $r_{rq} = 0.9$  the program found over 60 different implicit equations of linear regression with  $r > r_{rq}$ , almost impossible to obtain by hand or in some program with statistics kernel. If value of  $r_{rq}$  is increased to  $r_{rq} = 0.99$ , number of implicit equation founded is reduced at 29. For  $r_{rq} = 0.999$  number of implicit equation founded is 12.

That is a large set! If we are interested to study dependence between two variables form set, then we select the proper table from output of the program. Best result is displayed in table 4, and it correlate the temperatures  $T_1$  and  $T_M$ :

Table 4. Linear regression between martensitic, premartensitic and intermartensitic temperatures

$x_0$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	Equation	Residue	Correlation
0	0	0	0	0	1	1	$+x_5*1.00-x_6*1.80=-2.73*10^2$	0.37	0.98369
0	0	0	0	0	1	1	$-x_5*0.55+x_6*1.00=+1.50*10^2$	0.21	0.98369
0	0	0	0	0	1	1	$+x_5*3.66*10^{-3}-x_6*6.62*10^{-3}=-1.00$	0.42	0.98369

If we are looking for totally dependent variables (and here exists, sum of concentrations is 100%), the program finds it and also eliminate one of them from set. In table 5 is showed program response for correlating variables  $x_2$ ,  $x_3$  and  $x_4$  (dependent variable:  $x_2$ ):

Table 5. Founded dependent variable in group of concentrations of Ni( $x_2$ ), Mn( $x_3$ ) and Ga( $x_4$ )

$x_0$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	Equation	Residue	Correlation
0	0	1	1	1	0	0	$+x_2*1.00+x_3*1.00+x_4*1.00=+1.00*10^2$	0.00	1.00000
0	0	1	1	1	0	0	$+x_2*1.00+x_3*1.00+x_4*1.00=+1.00*10^2$	0.00	1.00000
0	0	1	1	1	0	0	$+x_2*1.00+x_3*1.00+x_4*1.00=+1.00*10^2$	0.00	1.00000
0	0	1	1	1	0	0	$-x_2*1.00*10^{-2}-x_3*1.00*10^{-2}-x_4*1.00*10^{-2}=-1.00$	0.00	1.00000

If we are looking for dependences between  $e/a$  and concentrations, simply select the founded equations from program output. In table 6 is showed the dependence of  $e/a$  by concentration of Ni and Mn:

Table 6. Dependence of  $e/a$  by Ni( $x_2$ ) and Mn( $x_3$ ) expressed by explicit and implicit equations

$x_0$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	Equation	Residue	Correlation
0	1	1	1	0	0	0	$+x_1*1.00-x_2*7.02*10^{-2}-x_3*3.99*10^{-2}=+2.98$	$7.35*10^{-4}$	0.99995
0	1	1	1	0	0	0	$-x_1*1.42*10^1+x_2*1.00+x_3*0.56=-4.25*10^1$	$1.55*10^{-3}$	0.99997
0	1	1	1	0	0	0	$-x_1*2.50*10^1+x_2*1.75+x_3*1.00=-7.47*10^1$	$5.43*10^{-3}$	0.99992
0	1	1	1	0	0	0	$-x_1*0.33+x_2*2.34*10^{-2}+x_3*1.33*10^{-2}=-1.00$	$1.86*10^{-3}$	0.99995

If we are looking to express one of the temperatures depending by concentrations, then the following equations are useful (table 7).

The equations that contain maximum of independent terms (without one concentration) is given at the end of output file of the program (table 8).

Table 7. Dependence of  $T_M$  (premartensitic temperature) by Mn( $x_3$ ) and Ga( $x_4$ ) concentrations

$x_0$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	Equation	Residue	Correlation
0	0	0	1	1	0	1	$+x_3*0.56+x_4*1.00+x_6*2.36*10^{-2}=+4.45*10^1$	$4.88*10^{-2}$	0.99300
0	0	0	1	1	0	1	$+x_3*2.37*10^1+x_4*4.23*10^1+x_6*1.00=+1.88*10^3$	0.16	0.99012

Table 8. Most comprehensive multi-linear dependence in data set

$x_0$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	Equation	Residue	Correlation
1	1	0	1	1	1	1	$-x_0*6.77*10^{-4}+x_1*1.00+x_3*2.79*10^{-2}+x_4*6.61*10^{-2}+x_5*5.27*10^{-6}-x_6*1.09*10^{-4}=+9.82$	$4.12*10^{-4}$	0.99999
1	1	0	1	1	1	1	$-x_0*2.42*10^{-2}+x_1*3.57*10^1+x_3*1.00+x_4*2.36+x_5*1.88*10^{-4}-x_6*3.92*10^{-3}=+3.51*10^2$	$4.36*10^{-3}$	0.99995
1	1	0	1	1	1	1	$-x_0*1.02*10^{-2}+x_1*1.51*10^1+x_3*0.42+x_4*1.00+x_5*7.97*10^{-5}-x_6*1.65*10^{-3}=+1.48*10^2$	$1.98*10^{-3}$	0.99999
1	1	0	1	1	1	1	$+x_0*6.17-x_1*9.11*10^3-x_3*2.54*10^2-x_4*6.03*10^2-x_5*4.80*10^{-2}+x_6*1.00=-8.95*10^4$	$9.44*10^{-2}$	0.99668

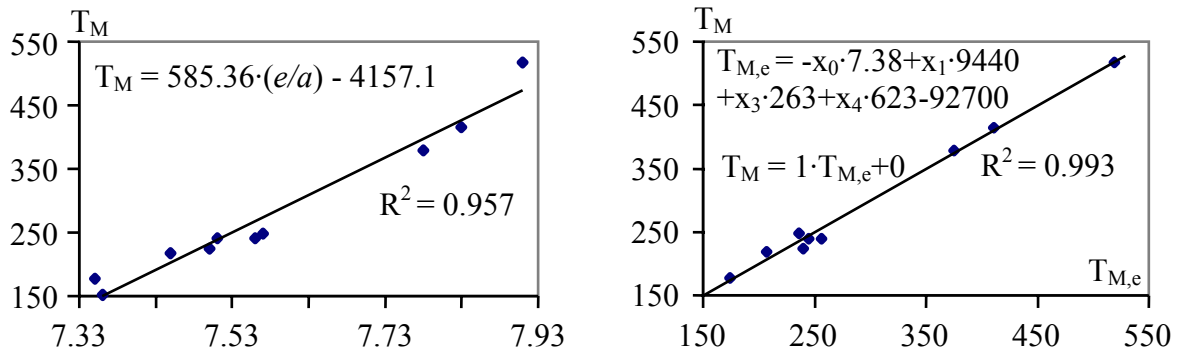


Fig. 3. (a) Regression between  $T_M$  and  $e/a$ ; (b)  $T_M$  and (Alloy state,  $e/a$ , %Mn, %Ga)

In figure 3 are plotted some selected dependences from data set. Figure 3a show a mono-variable dependence between  $T_M$  and  $e/a$ , and figure 3b show a multi-linear variable dependence between  $T_M$  and alloy state (codified by -1 and 1) electron/atom ratio and composition (%Mn and %Ga).

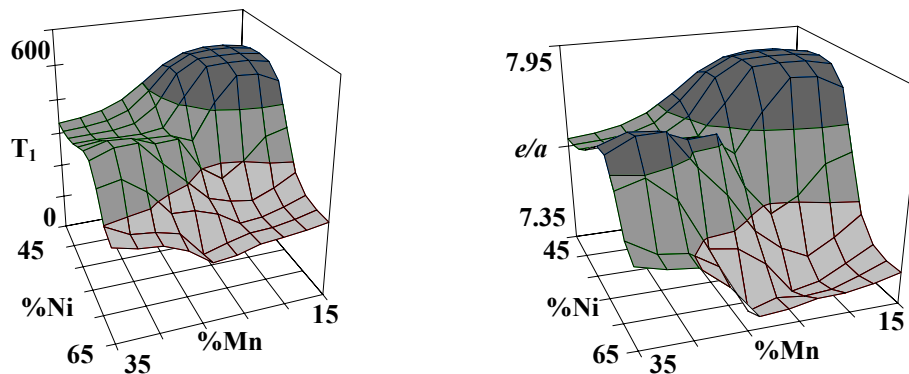


Fig. 4. (a) surface plot of  $T_1$  and (b)  $e/a$  by composition (%Ni, %Mn)

Figures 4b and 4c show surface dependences of  $T_1$  and respectively  $e/a$  of composition (%Ni, %Mn).

## 5. CONCLUSIONS

Looking to the output sums of residues from tables, is easy to observe now that the properties: type of alloy, and his martensitic, intermartensitic and premartenistic temperatures are interrelated together; these properties have the same order of sum residues in global equation, that is also expected conclusion. Very small same order of sum residues for concentrations suggest a strong interrelation between them, that is also true, because  $\%Ni + \%Mn + \%Ga = 100$ . This conclusion lead to consider the 3D plots fitted in figure 3 (b and c) of electron/atom ratio and  $T_1$  temperature dependencies by concentration (%Ni,%Mn). The figure 3a prove good correlation between  $T_1$  and  $e/a$ .

## 6. ACKNOWLEDGMENTS

Author is grateful to the rector of Technical University of Cluj-Napoca, prof. dr. eng. Gheorghe Lazea for his policy on promoting information technology and also to the university staff for support related to the internet connection of comp web server.

Useful support was also benefit from Romanian Ministry of Education for finance funding of MEC/CNCSIS contract 468/"A"/2002 and 281/"A"/2002.

## REFERENCES

- [1] Nascu H., Jäntschi L., Hodisan T., Cimpoiu C., Câmpan G., Rev. Anal. Chem., 1999; XVIII-6:409.
- [2] Jäntschi L., Unguresan M., In: Proc. of the Int. Conf. Qual. Contr. Aut. Rob., Cluj-Napoca, Romania: Technical University of Cluj-Napoca, 2002; 1:194.
- [3] Jäntschi L., Nascu H. I., In: Proc. of the Int. Conf. Qual. Contr. Aut. Rob., Cluj-Napoca, Romania: Technical University of Cluj-Napoca, 2002; 1:259.
- [4] Jäntschi L., Unguresan M., Oradea Univ. Annals, Math. Fasc., 2001:VIII:105.
- [5] Marta L., Deac I., Fruth I., Zaharescu M., Jäntschi L., In : Proc. of the Int. Conf. Sci. Mater., Cluj-Napoca, Romania: Technical University of Cluj-Napoca, 2000; 2:627.
- [6] Marta L., Deac I., Fruth I., Zaharescu M., Jäntschi L., In: Proc. of the Int. Conf. Sci. Mater., Cluj-Napoca, Romania: Technical University of Cluj-Napoca, 2000; 2:633.
- [7] Cimpoiu C., Jäntschi L., Hodisan T., J. Plan. Chromatog. - Modern TLC, 1998; 11:191.
- [8] Cimpoiu C., Jäntschi L., Hodisan T., J. Liq. Chromatogr. Rel. Techn., 1999; 10:1429.
- [9] Diudea M., Jäntschi L., Graovac A., In Proc. of the Math/Chem/Comp'98 - Dubrovnik Int. Cour. & Conf. Interf. Math., Chem. Comp. Sci., Dubrovnik, Yugoslavia, 1998; 1:3.
- [10] Jäntschi L., Katona G., Diudea M., Commun. Math. Comput. Chem. (MATCH), 2000; 41:151.
- [11] Sârbu C., Jäntschi L., Revista de Chimie, 1998; 1:19.
- [12] Diudea M., Gutman I., Jäntschi L., Molecular Topology, Commack, New York: Nova Science, 2001. p. 212.
- [13] Chernenko V.A., Pons J., Seguy' C., Cesari E., Acta Mater., 2002; 50:53.