Annals of West University of Timisoara,

Series Chemistry **12 (3)** (2003) 1137-1144

CASE STUDY OF COALS FROM FĂRCĂŞEŞTI AREA. 2. DEPENDENCIES INCLUDING FIXED CARBON DETERMINATIONS

Elena Maria PICĂ[a*], Lorentz JÄNTSCHI[a]

[a] Technical University of Cluj-Napoca, Faculty of Science and Engineering of Materials, Chemistry Department, Muncii Street, 101-103, Cluj, 3400, ROMANIA

[*] empica@yahoo.com

SUMMARY

A statistical study of analysis results was made for lignite from the Fărcăşeşti area (Gorj County, Romania), exemplified for the nine characteristic properties, as moisture content (imbibitions and hygroscopic), volatile, density, fixed carbon, sulfur, ash softening content, higher heating value and seam. Previously, the properties dependencies were investigate in pairs of two. In present study the properties was investigated using an automat processing routine for multivariate regression, available at address: http://academicdirect.ro/virtual_library/applied_statistics/ linear_regression/multiple/v1.5/

The program is capable to identify multiple dependencies between given properties. Few significant results were obtained, that make possible to predict higher heating value Q_{si} from fixed carbon content C_{fi} and other measured properties and make possible reducing number of determinations. Present article is focused on identifying dependencies between W_{ii}, W_{hi}, V_i, C_{fi}, S_{ti}, Q_{si}, t_i, r_o and seam (see text). Application of the model among others at prospecting new coalfields and coal conversion, can contribute to the reduction of drilling and analysis costs.

Keywords: coal analysis, regression models, software implementation study

INTRODUCTION

Many statistical procedures for processing data are 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]. Modeling of structure is benefit to property predictions [6, 7]. Nonstandard statistical evaluation procedures then are helpful [8].

The design of statistical processing program is depicted in fig. 1:

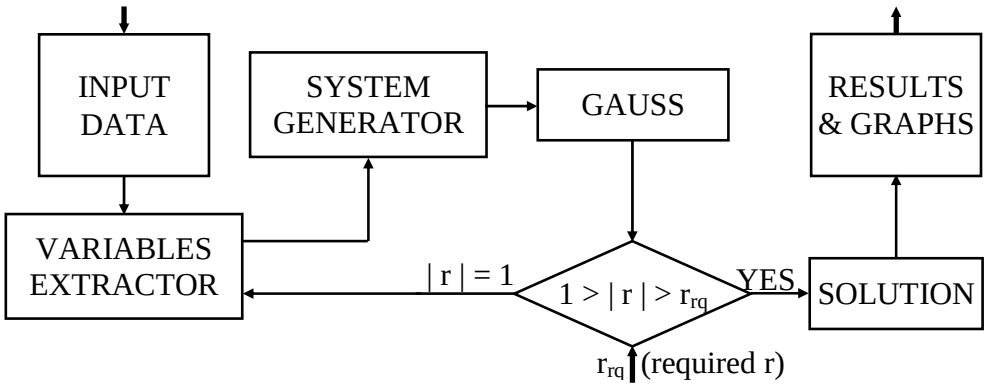

Figure 1. Program design scheme

The INPUT module read a text format data, process input data, split it into rows and columns. The rows represent the values of variables and columns represent the variables. The INPUT module passes data in matrix form to the VARIABLES EXTRACTOR module. This module starts with pairs of columns from data matrix; continue with sets of three columns and so on, until entire set is browsed. For every set, pass it to the SYSTEM GENERATOR. Another input in VARIABLES EXTRACTOR module comes from DECISION module. If the DECISION module detects a correlation of absolute coefficient 1 then it passes to the VARIABLES EXTRACTOR the first variable

label from the regression equation and for the next extractions of variables the passed variable is marked as dependent and is not passed never again to the SYSTEM GENERATOR module. The SYSTEM GENERATOR computes average means, needed by GAUSS module. 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 SYSTEM GENERATOR is computed by formulas (equation 1):

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

Linear regression and PLS (partial least squares) are most used methods in statistical processing of data. Presented method (SYSTEM GENERATOR and GAUSS modules) uses them. GAUSS module solves a linear system of equations. 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 (equation 2):

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

When system is solved a unique solution is found. Then, system extended matrix contain at column n_cols the coefficients of regression equation (equation 3):

$$a_1 \cdot x_1 + ... + a_i \cdot x_i + ... + a_{n\_cols} \cdot x_{n\_cols} + a_{n\_cols+1} = 0,\ a_{n\_cols+1} = -1$$

where the coefficients a_{n_cols+1} and a_{n_cols+1} are resulted regression coefficients. Note that equation (3) is in implicit form; to obtain an explicit form is necessary to extract dependent variable from (3). The last coefficient is assigned to -1. At the end of module SOLUTION it result an implicit linear regression equation between given variables through his values in columns (equation 3). Equation 3 can be exploited to obtain explicit linear regression equations for each variable which has no null coefficient a_i (equation 4):

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

A graphical interface was built in PHP with a TEXTAREA for input data and an INPUT SUBMIT button for submitting data to the server. The server is a FreeBSD Unix based server (5.0 DP1 software version) with an Apache web server (1.3.26 software version) running on.

The server is hosted in educational network of Technical University of Cluj-Napoca with address 193.226.7.211 and name academicdirect.ro. The PHP language was compiled with GDI (graphical device interface) and MySQL (database server) support and the PHP software version is 4.2.3. The MySQL database server is also installed and running on and his software version is 3.23.52. The input interface is presented in figure 2:

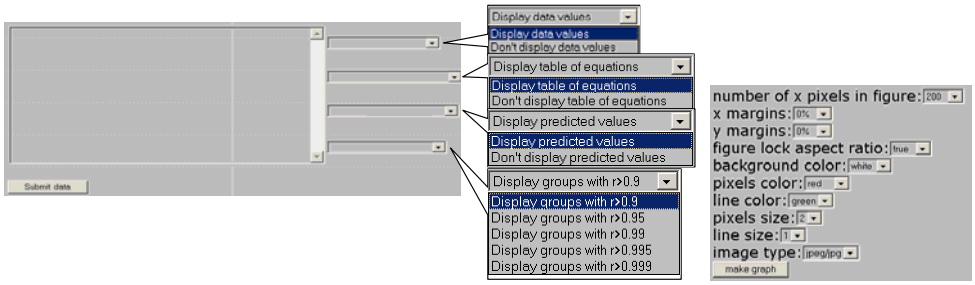

Figure 2. The output of http://193.226.7.211/~lori/research/regression/linear/v1.5/index.php and the output of http://193.226.7.211/~lori/research/regression/graphics/lin_plot/do.php

The output of the RESULTS&GRAPHS module is passed depending on user equation selection to another PHP program that makes graphical representation of regression [9].

DATA MINING

To characterize a coal seam the results of proximate analyses as function of depth as related to the initial sample (i), the sample for analysis (a), or anhydrous sample (anh) can be considered [10, 11].

A set of measured data from Farcăşeşti area was taken into statistical analysis. The probes for analysis were taken from the 64040.15 platform at different seams. The analysis results are given in table 1. All measured data from table 1 refer to the initial sample “i” and

are expressed in percents (excepting the ash softening temperature, the density and the number of seam).

Table 1. Data values and Q_{si} predicted values

W_{ii}	W_{hi}	V_i	C_{fi}	S_{fi}	Q_{si}	t_i	r_o	seam	Q_{si} predicted values (see text)			
x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$f(V_i, C_{fi})$	$f(W_{ii}, V_i, C_{fi})$	$f(V_i, C_{fi}, S_{ti})$	$f(W_{ii}, W_{hi}, V_i, C_{fi}, S_{ti})$
A	B	C	D	E	F	G	H	I				
34.4	8.1	23.4	14.7	1.03	2225	1130	1.23	16	2164	2263	2176	2261
24.7	9.9	25.3	17.5	0.69	2529	1100	1.19	14	2531	2473	2560	2522
27.1	7.9	20.1	12.1	0.64	1802	1250	1.39	13	1708	1737	1742	1759
25.0	10.5	26.2	19.2	0.90	2700	1250	1.22	12	2735	2663	2756	2714
29.0	8.4	22.1	14.6	0.95	2117	1150	1.11	10	2058	2087	2078	2091
33.0	9.6	26.0	17.9	1.25	2641	1105	1.10	10	2616	2663	2617	2691
28.5	9.1	25.5	18.2	1.32	2590	1120	1.25	10	2603	2587	2603	2559
32.2	8.9	27.2	18.6	0.97	2816	1130	1.12	10	2762	2791	2775	2769
33.3	9.6	25.0	18.6	0.93	2647	1115	1.10	10	2597	2640	2619	2667
25.5	9.4	30.5	20.9	2.10	3043	1100	1.22	10	3194	3112	3148	3020
30.0	8.3	21.2	14.0	1.67	2025	1105	1.13	8	1943	1990	1929	1975
34.7	9.8	26.6	20.9	0.88	2919	1115	1.08	8	2902	2938	2927	2945
26.9	10.7	28.5	19.8	1.63	2983	1100	1.02	8	2956	2903	2936	2937
33.1	9.5	27.1	20.6	1.07	2949	1085	1.20	7	2915	2934	2928	2915
34.4	9.1	26.3	18.7	1.97	2692	1125	1.28	6	2703	2759	2671	2718
25.0	8.8	29.5	15.7	0.82	2650	1110	1.20	5.9	2700	2669	2707	2678
27.5	10.2	26.9	19.0	1.69	2737	1120	1.13	5.1	2772	2735	2752	2749
25.4	10.7	27.5	19.1	2.23	2741	1115	1.20	5	2825	2760	2778	2778

The imbibitions moisture content is W_{ii}, the hygroscopic moisture content is W_{hi}, the volatile content is V_i, the fixed carbon content is C_{fi}, the content of total sulfurs is S_{ti}, the higher heating value is Q_{si}, the ash softening temperature is t_i, the density is r_o, and the seam is represented by a number.

The used functions for Q_{si} prediction are:

$$f(x_2,x_3) = x_2 \cdot 74.8 + x_3 \cdot 80.5 - 770;\ r = 0.992;\ s = 0.098$$

$$f(x_0,x_2,x_3) = x_0 \cdot 13.1 + x_2 \cdot 75.1 + x_3 \cdot 69.4 - 9750;\ r = 0.994;\ s = 0.066$$

$$f(x_2,x_3,x_4) = x_2 \cdot 71.8 + x_3 \cdot 82.9 - x_4 \cdot 47.7 - 6740;\ r = 0.994;\ s = 0.079$$

$$f(x_0,x_1,x_2,x_3,x_4) = x_0 \cdot 16.3 + x_1 \cdot 65.6 + x_2 \cdot 76.2 + x_3 \cdot 51.3 - x_4 \cdot 37.2 - 1330;\ r = 0.996;\ s = 0.054$$

RESULTS AND DISCUSSION

Four plots are in figure 3 and represent the regressions between Q_{si} and calculated values (f functions) by a linear equation.

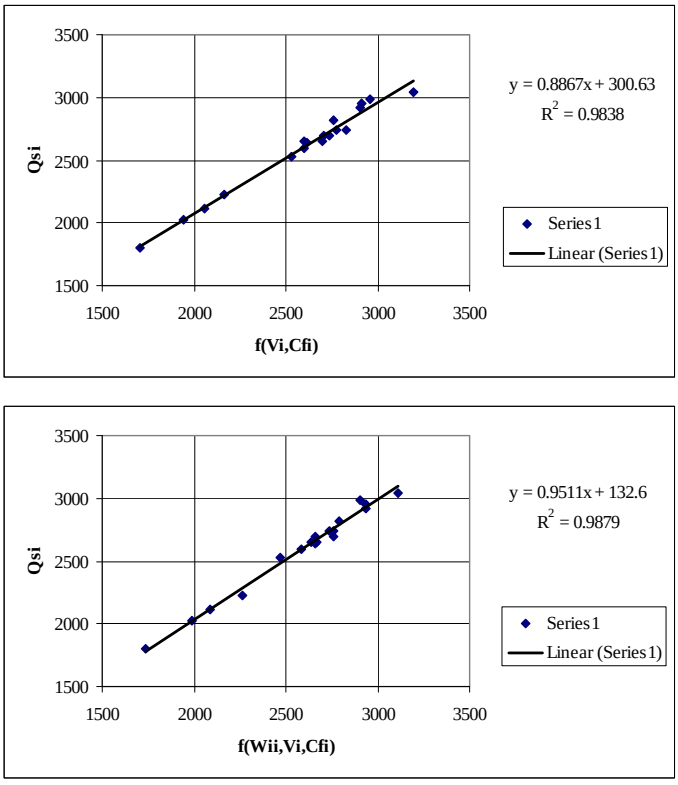

Figure 3. Dependencies of Q_{si} by other measured data from table 1

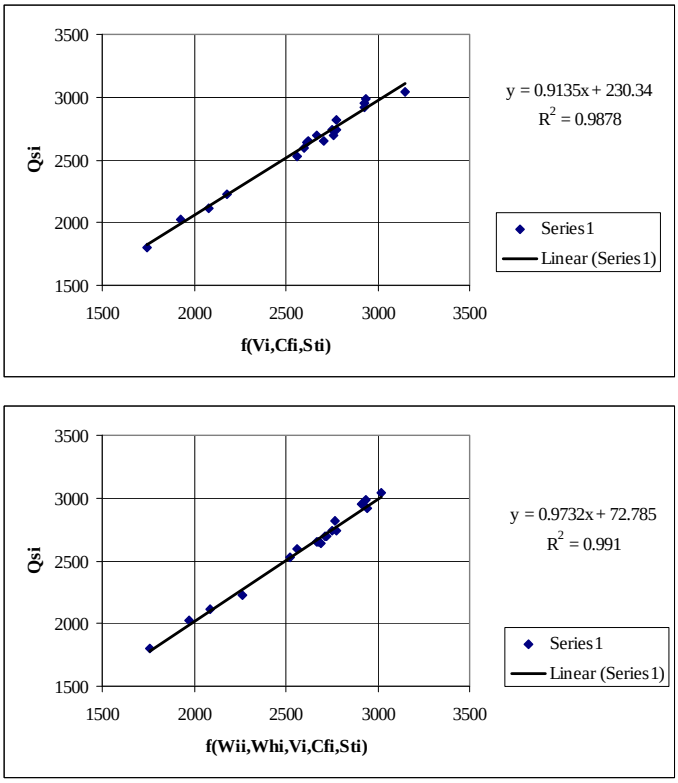

Figure 3. Dependencies of Q_{si} by other measured data from table 1 – continuing

First plot consider only dependence of higher heating value Q_{si} by the fixed carbon content C_{fi} and the volatile content V_i and predict them with an squared r about 0.98.

Following dependencies prove that including more variables in regression do not add more accuracy in prediction.

Dependency functions of Q_{si} include fixed carbon C_{fi} parameter, as we expected and is better correlation than the fixed carbon is not included [12].

CONCLUSIONS

The study shows the possibility of reducing number of analysis for physical and chemical parameters of coals without reducing the quality of information. Are obtained the dependency of higher heating value Q_{si} by the fixed carbon content C_{fi} and the volatile content V_i and predict them with an squared r about 0.98.

The original program used was making possible to investigate all correlations between measured data, and simplify the research study of correlations.

REFERENCES

1. H. Nascu, L. Jäntschi, T. Hodisan, C. Cimpoiu, G. Câmpan, Some Applications of Statistics in Analytical Chemistry, Reviews in Analytical Chemistry, Freud Publishing House, XVIII(6), 1999, 409-456.
2. L. Jäntschi, M. Unguresan, MathLab. Statistical Applications, International Conference on Quality Control, Automation and Robotics, May 23-25 2002, Cluj-Napoca, Romania, work published in volume "AQTR Theta 13" (ISBN 973-9357-10-3) at p. 194-199.
3. L. Jäntschi, H. Nascu, Numerical Description of Titration (MathCad application), International Conference on Quality Control, Automation and Robotics, May 23-25 2002, Cluj-Napoca, Romania, work published in volume "AQTR Theta 13" (ISBN 973-9357-10-3) at p. 259-262.
4. L. Jäntschi, M. Unguresan, Parallel processing of data. C++ Applications, Oradea University Annals, Mathematics Fascicle, 2001, VIII, 105-112, ISSN 1221-1265.
5. L. Marta, I. G. Deac, I. Fruth, M. Zaharescu, L. Jäntschi, Superconducting Materials: Comparision Between Coprecipitation and Solid State Phase Preparation, International Conference on Science of Materials, August 20-22, 2000, Cluj-Napoca, Romania, work published in volume 2 (ISBN 973-686-058-2) at p. 627-632.
6. M. Diudea, L. Jäntschi, A. Graovac, Cluj Fragmental Indices, Math/Chem/Comp'98 (The Thirteenth Dubrovnik International Course & Conference on the Interfaces among Mathematics, Chemistry and Computer Sciences), June 22-27, 1998.
7. L. Jäntschi, G. Katona, M. Diudea, Modeling Molecular Properties by Cluj Indices, Commun. Math. Comput. Chem. (MATCH), 2000, 41, 151-188, ISSN 0340-6253.
8. C. Sârbu, L. Jäntschi, Statistic Validation and Evaluation of Analytical Methods by Comparative Studies. I. Validation of Analytical Methods using Regression Analysis (in Romanian), Revista de Chimie, Bucuresti, 49(1), 1998, 19-24, ISSN 0034-7752.
9. L. Jäntschi, Free Software Development. 5. Statistical regressions plotting with GDI interface on Apache web server and PHP module, Leonardo Electronic Journal of Practices and Technologies, Cluj-Napoca, in press.
10. Blum I., Barca Fr., Chemistry and preparation of solid combustors (in Romanian), Ed. Did. Ped., Bucharest, 1966.
11. Enache C-tin., Focşa I., Pică Elena Maria, Niac G., The Study of Lignite from the Fărcăşeşti Area, Acta Technica Napocensis, Section: Construction Machines and Materials, 36 (1993), pp.79-86.
12. Pică E. M., Jäntschi L., Case Study of Coals from Fărcăşeşti Area. 1. Dependencies Including Fixed Carbon Determinations, RomPhysChem11, submitted.