DOI: 10.2478/ctg-2014-0010

Marcin Sajdak[1], Łukasz Smędowski[2]

[1]Institute for Chemical Processing of Coal,
Zamkowa Str. 1, 41-803 Zabrze.
E-mail: msajdak@ichpw.zabrze.pl
[2]Institute for Chemical Processing of Coal,
Zamkowa Str. 1, 41-803 Zabrze.
E-mail: lsmedowski@ichpw.zabrze.pl

Application of multivariate data analysis in the construction of predictive model for the chemical properties of coke

Key words:
coke, chemical composition,
prediction,
mathematical model.

Abstract

The aim of this work was to develop a statistical model which can predict values describing chemical composition of cokes performed in industrial scale. This model was developed on the basis of data that were taken from the production system used in the one of Polish coking plant. Elaborated equation include quality parameters of initial coals that form coal blends as well as contribution of additions such as coke and petrochemical coke. These equations allow to predict chemical composition of coke, e.g. contributions of: sulphur, ash, phosphorus and chlorine within the coke. A model was elaborated with use of STATISTICA 10 program and it is based on factor and multiply regression analyses. These analyses were chosen from among few kinds of regression analyses. They allowed to develop prediction model with the required goodness of fit between calculated and actual values. Goodness of fit was elaborated with:

- residuals analyses,
- residues normality and predicted normality
- mean absolute error
- Pearson correlation confidence

DE GRUYTER OPEN

Introduction

Cokes manufactured at high temperatures are used in the steel industry for the reduction of iron ore to pig iron in blast furnaces. Fabrication of such metallurgical cokes, needs a large amount of very good coking coals because the very good quality of this product is required (Hereźniak & Warzecha, 2009). From among a lot of chemical parameters the most significant are: ash, sulphur and phosphorus contents. A mineral matter that is contained in coke cause decreasing of its heating value, reactivity and mechanical strength. Presence of ash also cause increasing of energy demand for blast furnace process and the effect is that the consumption of coke is increasing, e.g. 1% increase of ash content causes the 2,5% increase of the unit coke consumption (Karcz, 1991). Sulphur that is contained in the coke represents ca. 70 – 90% of all sulphur that is enforced to a blast furnace by all of raw materials used in the process. It is known that the 0,1% increase of sulphur content in the coke cause that its consumption for one tone of pig iron increase about 0,3 – 1,1 % and the productivity of the blast furnace decrease about 2,0 % (Karcz, 1991). The presence of phosphorus in the coke strongly deteriorates the quality of pig iron and consequently, the resultant steel becomes brittle (Zieliński, 1986). Hence, it can be assumed that the possibility of the prediction of the chemical properties of coke basing on the results of chemical composition of initial coals analyses is very important topic. The application of predictive model could indicate how blending coking coals to obtain coke with desired chemical properties. The aim of this work was to develop a statistical model which can predict values describing chemical composition of cokes performed in industrial scale.

Input data

Data that describe chemical composition of initial coals and resultant cokes were taken from the production system used in one of the Polish coking plant. The yield of analyzed cokes, e.g. the ratio of mass of coke to the mass of carbonized blend, was in the same level for all samples studied, so it can be assumed that this parameter does not influence on the chemical composition of cokes.

Statistical methods

To this research we have chosen few regression methods (Friedl et al. 2005, Lei et al. 2011), which were used to build models that estimate chemical properties of coke. Such prepared set of data was input matrix for program STATISTICA 10 (StatSoft, 2011). During the analysis matrix was divided into two groups: teaching and testing in 70:30 proportion. Data from both groups were analysed using following methods of regression (Abnisa et al. 2011, Parikha et al. 2005):

- K – Nearest Neighbour Model, KNN,
- Support vector machine - SVM Model,
- Automated Neural Networks - SANN Model,
- Tree-Building Algorithm - CHAID Model,
- Tree-Building Algorithm - C&RT Model,
- Regression Model, REG,
- General Regression Model, GRM.

Some above mentioned methods (C&RT, regression method) have also been successfully used to build predictive models of biomass heat combustion and relationship between elements in bio-char (Sajdak, 2013, Sajdak et al 2013). All models obtained during our research were analysed in terms of goodness of fit to the actual data (Zhang et al. 2004, Álvarez et al. 2007). The created prediction models, which were the result of the above-mentioned methods, were tested using data that was not included in the predictive model (test set). Comparability between the model and the real data was measured using 10 times cross validation, uncertainty and misfit errors were collected. Values for each of statistic indexes were calculated using equations shown below (Chun-Yang, 2011, Dĩez et al. 2002):

Mean Squared Error (MSE)

(1)

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(\overline{X}-x_i\right)^2$$

where:

$\overline{X}$ - vector of n predictions values

x_i - true values

Mean Absolute Error (MAE)

(2)

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|X_i-x_i|=\frac{1}{n}\sum_{i=1}^{n}|e_i|$$

where:

e_i -absolute error

X_i -predictions values

Standard deviation of mean values

(3)

$$\sigma_{\overline{X}}=\frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}}{\sqrt{N}}$$

where:

$\overline{x}$ - mean values

Relative Standard Deviation RSD

(4)

$$RSD=\frac{\sigma}{\overline{x}_i}=\frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}}{\overline{x}_i}$$

Pearson correlation r

(5)

$$r_{xy}=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\left(n-1\right)s_x s_y}$$

Where:

$\overline{x}$ and $\overline{y}$ - the sample means of X and Y,

s_x and s_y - the sample standard deviations of X and Y.

Data set in form of the matrix consisted of 13 independent variables, e.g. chemical properties of coals from various coal mines, and their contribution in the coal blends were used to performed analysis. Chemical properties of cokes, e.g. ash, sulphur, chlorine and phosphorous contents, produced on the base of coal blends were used as a dependent variables. Maximum numbers of data in each matrix is about 7000.

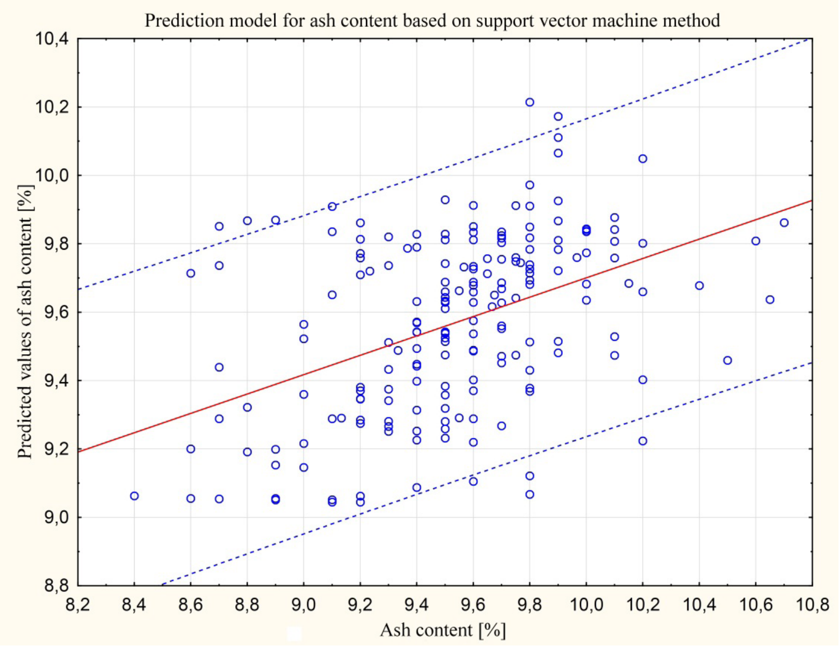
Results and discussion

Each statistic operation described in previous chapter of this paper was carried out in STATISTICA software (StatSoft, 2012). The results of quality of data analysis performed with use of the various regression methods has been presented in the table 1. It can be seen that the most universal method, which can be used to build models for estimation of coke properties, is general regression method, GRM. In few tasks, methods such as: support vector machine method, automated neural networks method and tree building algorithm proved to be better than general regression method. We had to deal with this case in the ash content estimation, where the best method to ash prediction is support vector machine method (SVM). In this case Pearson's correlation coefficient is about 48% and mean absolute error values is about 0,27. Figure 1 presents a scatter plot of actual and predicted values of ash content in tested materials. Dotted lines in figure 1 represent the prediction interval for 95%. Prediction interval informs about the single forecasts of the dependent variable. Range prediction of the dependent variable is the range of values, within which, with a proper probability, we expect the next observation for the given values of the independent variables. It can be seen in figure 1 almost all points lie inside the prediction intervals, which is equal to +-6,4% for ash content. In this range, includes expanded uncertainty value of ash content analysis, which is equal to 2,7%

Table 1. Goodness of fit values of created prediction models for determinations of sulphur, phosphorus, chlorine and ash content in tested materials.

Analysis type	Mean squared error	Mean absolute error	Standard deviation of mean values	Relative standard deviation	Person's correlation coefficient
Sulphur					
KNN	0,0031	0,0397	0,0098	0,0688	0,6720
SVM	**0,0018**	**0,0315**	**0,0056**	**0,0541**	**0,7831**
SANN	0,0019	0,0335	0,0058	0,0570	0,7631
CHAID	0,0020	0,0342	0,0057	0,0581	0,7402
CART	0,0027	0,0367	0,0081	0,0628	0,6813
REG	**0,0017**	**0,0314**	**0,0052**	**0,0536**	**0,7843**
GLM	**0,0017**	**0,0314**	**0,0052**	**0,0536**	**0,7843**
Phosphorus					
KNN	0,0001	0,0070	0,0263	0,1173	0,1481
SVM	0,0001	0,0064	0,0252	0,1062	0,1455
SANN	0,0001	0,0060	0,0232	0,0983	0,2283
CHAID	**0,0001**	**0,0053**	**0,0208**	**0,0891**	**0,3987**
CART	0,0001	0,0076	0,0294	0,1242	0,2072
REG	**0,0001**	**0,0054**	**0,0188**	**0,0903**	**0,4390**
GLM	**0,0001**	**0,0054**	**0,0188**	**0,0903**	**0,4390**
Chlorine					
KNN	0,0001	0,0094	0,0624	0,2085	0,2822
SVM	0,0001	0,0081	0,0529	0,1885	0,1951
SANN	**0,0001**	**0,0056**	**0,0280**	**0,1319**	**0,6113**
CART	0,0001	0,0081	0,0470	0,1771	0,2960
REG	**0,0001**	**0,0062**	**0,0302**	**0,1435**	**0,6175**
GLM	**0,0001**	**0,0062**	**0,0302**	**0,1435**	**0,6175**
	Ash				
KNN	0,1834	0,3136	0,0021	0,0330	0,4019
SVM	**0,1344**	**0,2711**	**0,0015**	**0,0286**	**0,4831**
SANN	0,1512	0,2870	0,0017	0,0304	0,4407
CHAID	0,1450	0,2841	0,0017	0,0302	0,4106
CART	0,2203	0,3526	0,0025	0,0371	0,3517
REG	0,1519	0,3006	0,0017	0,0319	0,3773
GLM	0,1519	0,3006	0,0017	0,0319	0,3773

Figure 1. The scatter plot of real and predictive values of ash content in tested materials.



In Figure 2 there are included prediction charts for the rest of independent variables studied, e.g. sulphur (Fig. 2a), phosphorous (Fig. 2b) and chlorine (Fig. 2c) contents. It can be seen in the presented charts that almost all points lie inside the prediction intervals, which is the same for all independent variables and are equals 95%. Differences between real values and predicted values aren't too big because the maximum range is about 10% for sulphur, 8,8% for chlorine, 6,4% for ash content and 13,3% for phosphorus content. This values isn't too big because this range, includes an expanded uncertainty value for sulphur content analysis, which is equal to 3,7%, for chlorine content is equal to 16,6% and for phosphorus content is equal to 8,6%. Only in chlorine case expanded uncertainty value is twice as large and is equal to 16,6%. For further analysis, selected models based on general regression method, these models have been validated on new data set. As the result mathematical equations describing the properties of coke were created:

(6)

$$C_{X_{coke}} = \sum_{i=1}^{n} \left(\mu_i \cdot C_{X_i} \cdot u_i \right) + const$$

where:
C_{Xcoke} – prediction values of analysed variables,
μ_i – constant coefficients,
C_{Xi} – content of analysed variables in coal from i – coal mine,
u_i – mass fraction of coal from i – coal mine in the blend.

The constant coefficients of each values of analysed variables has been presented in the table 2.

Figure 2. The scatter plot of real and predictive values of a) sulphur, b) phosphorus, c) chlorine content in tested materials.

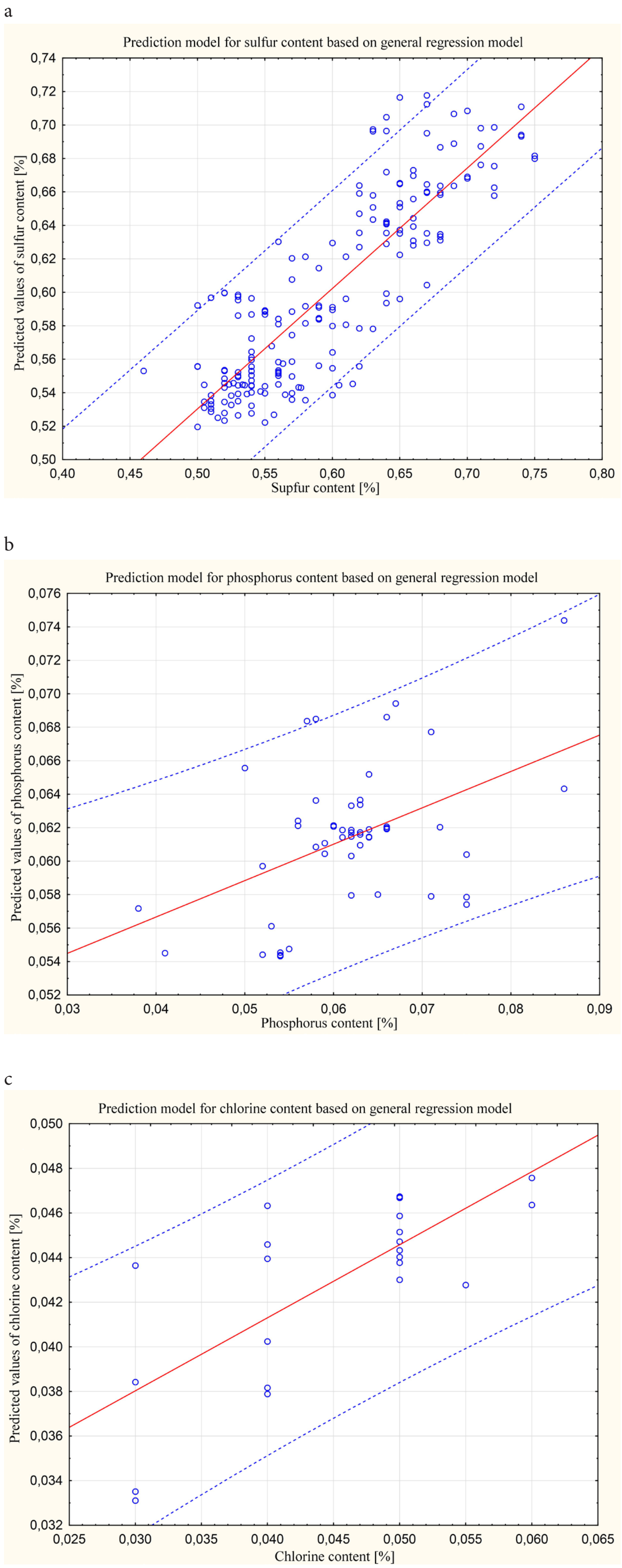


Table 2. Values of constant coefficients of analysed variables.

Variables		S [%]	P [%]	Cl [%]	Ash [%]
const.		0,616	0,048	0,102	8,255
Coal mine	BN	0,078	0,108	-0,013	0,150
	PN	-0,453	0,862	-0,459	0,154
	Z	-0,077	0,238	-0,906	0,242
	DK	0,745	-58,835	0,000	0,000
	JM	-0,681	-7,582	-1,063	0,266
	S/KN	0,210	0,000	0,000	-0,140
	CSM	-0,400	1,060	-2,861	-0,090
	PK	0,030	0,905	-1,626	0,309
	KL	0,087	-0,743	-2,093	0,057
	KX	1,199	0,000	0,000	-0,018
	PC	0,743	0,000	0,000	-0,409
	BD/KP	0,319	0,754	-0,374	0,310

Notation:
BN – Borynia coal mine,
PN – Pniówek coal mine,
Z – Zofiówka coal mine,
DK – Darkov coal mine,
JM – JasMos coal mine,
S/KN – Szczygłowice/Knurów coal mine,
CSM – CSM coal mine,
PK – Paskov coal mine,
KL – Columbian coal,
BD/KP – Budryk/Krupiński coal mine,
KX – coke,
PC – petroleum coke.

Conclusions

The research presented in this article was performed to verify which of the available methods of regression analysis can provide a predictive model for the determination, with good precision (accuracy), of the sulphur, phosphorus, chlorine and ash contents in metallurgical coke. During the analysis, it was determined that the created models gives the best results in terms of the prediction of ash content where have been used support vector machine method. On the other side sulphur, phosphorus and chlorine contents were predicted with the highest goodness of fit with use of general regression method. Differences between real values and predicted values have been calculated and were respectively: 10% for sulphur, 8,8% for chlorine, 6,4% for ash content and 13,3% for phosphorus content. Such models should be very useful for cokemakers, because they will be able to predict theirs coke quality and to optimize the consumption of very expensive high-quality coking coals in their plants.

Acknowledgements

This study was financed under the Smart Coking Plant R&D Project (Innovative Economy Programme (POIG), Contract no 01.01.02.-24-017/08).

References

Abnisa F., Wan Daud W.M.A., Sahu J.N. (2011), Optimization and characterization studies on bio-oil production from palm shell by pyrolysis using response surface methodology, *Biomass and Bioenergy Volume 35, Issue 8, August, Pages 3604–3616*

Álvarez R., Díez M.A., Barriocanal C., Díaz-Faes E., Cimadevilla J.L.G., (2007), An approach to blast furnace coke quality prediction, *Fuel*, 86, 14, 2159-2166,

Chun-Yang Yin, (2011), Prediction of higher heating values of biomass from proximate and ultimate analyses, *Fuel* 90 1128–1132

Dĩez M.A, Alvarez R, Barriocanal C, (2002), Coal for metallurgical coke production: predictions of coke quality and future requirements for cokemaking, *International Journal of Coal Geology, 50, 1–4, 389-412,*

Friedl A., Padouvas E., Rotter H., Varmuza K. (2005) Prediction of heating values of biomass fuel from elemental composition, *Analytica Chimica Acta 544 191–198*

Hereźniak, W., Warzecha, A. (2009) *Międzynarodowy rynek węgla koksowego i koksu – stan obecny i prognozy rozwoju, Karbo, 4, 197 - 206.*

Karcz, A., (1991) Koksownictwo, cz. 1, Wydawnictwo AGH, Kraków (in Polish).

Lei H., Ren S., Wanga L., Bu Q., Julson J., Holladay J., Ruan R., (2011), Microwave pyrolysis of distillers dried grain with solubles (DDGS)for biofuel production, *Bioresource Technology 102 6208–6213*

Parikha J., Channiwalab S.A., Ghosalc G.K., (2005), A correlation for calculating HHV from proximate analysis of solid fuels, *Fuel 84 487–494*

Sajdak M. (2013), Application of chemometrics to identifying solid fuels and their origin, *Cent. Eur. J. Chem., 11(2), 151-159*

Sajdak M. Piotrowski O. (2013), C&RT model application in classification of biomass for energy production and environmental protection, Cent. Eur. J. Chem., 11(2), 259-270

StatSoft, Inc. (2011) STATISTICA 10

Zhang Q., Wu X., Feng A., Shi M., (2004), Prediction of coke quality at Baosteel, *Fuel Processing Technology, 86, 1, 15, 1-11,*

Zieliński, H (eds.) (1986) Koksownitwo, Wydawnictwo Śląsk, Katowice (in Polish)

INNOWACYJNA
GOSPODARKA
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO