



ARCHIVES
of
FOUNDRY ENGINEERING

DOI: 10.1515/afe-2017-0102

Published quarterly as the organ of the Foundry Commission of the Polish Academy of Sciences

ISSN (2299-2944)
Volume 17
Issue 3/2017

117 – 122

Approximation of Ausferrite Content in the Compacted Graphite Iron with the Use of Combined Techniques of Data Mining

K. Regulski ^a, D. Wilk-Kołodziejczyk ^{a, b, *}, B. Kacprzyk ^c, G. Gumienny ^c, G. Rojek ^a, B. Mrzyglód ^a^a AGH University of Science and Technology, Faculty of Metals Engineering and Industrial Computer Science, Mickiewicza 30, Kraków, Poland^b Foundry Research Institute, ul. Zakopiańska 73, 30-418 Kraków, Poland^c Department of Materials Engineering and Production Systems, Lodz University of Technology, Stefanowskiego 1/15 Street, 90-924 Łódź, Poland

* Corresponding author: Email address: wilk.kolodziejczyk@gmail.com

Received 07.06.2017; accepted in revised form 28.07.2017

Abstract

This article presents the methodology for exploratory analysis of data from microstructural studies of compacted graphite iron to gain knowledge about the factors favouring the formation of ausferrite. The studies led to the development of rules to evaluate the content of ausferrite based on the chemical composition. Data mining methods have been used to generate regression models such as boosted trees, random forest, and piecewise regression models. The development of a stepwise regression modelling process on the iteratively limited sets enabled, on the one hand, the improvement of forecasting precision and, on the other, acquisition of deeper knowledge about the ausferrite formation. Repeated examination of the significance of the effect of various factors in different regression models has allowed identification of the most important variables influencing the ausferrite content in different ranges of the parameters variability.

Keywords: Application of information technologies in the field of foundry, Compacted graphite iron, Ausferrite, Data mining, Regression

1. Introduction

CGI (Compacted Graphite Iron), having a specific graphite form with large matrix contact surface, is a unique casting material. This type of cast iron tends to favour direct ferritization and is characterized by a complex of very interesting properties. Compared to grey iron, this material is characterized by higher strength properties and better ductility. Compared to spheroidal graphite iron, compacted graphite iron exhibits lower coefficient of thermal expansion, higher thermal conductivity, higher resistance to dynamic temperature changes, and better damping and casting properties. These advantages predestine the material

for a variety of uses. The first utilitarian use of this cast iron was for the high-speed brake discs. Currently it is mainly used for the construction of IC engine blocks, exhaust manifolds, etc. The complex properties of this type of cast iron are the subject of intensive research and many publications [1,2].

2. Microstructure of CGI

There are only few publications on the effect that alloying additives have on the microstructure and properties of compacted graphite iron [3,4]. By changing the chemical composition, it is

possible to modify to a large extent the cast iron microstructure and consequently its properties. There are numerous publications describing the possibility of obtaining ausferrite in compacted graphite iron, to produce next AVI (Austempered Vermicular Iron) or CAVI (Carbide Austempered Vermicular Iron) [5,6]. Ausferrite is a mixture of bainitic ferrite and carbon-saturated austenite. It is a desirable component of the microstructure, because under the influence of stress, as a result of twinning, this constituent undergoes martensitic transformation and thus strengthening. To obtain ausferrite, the casting should be subjected to heat treatment, which consists of quenching with isothermal holding within the range of austenite \rightarrow bainite transition temperature. There is also an alternative way to obtain ausferrite by modifying the cast iron chemical composition. This is done using molybdenum, copper, or nickel added in appropriate proportions.

2.1. Research methodology

Metal was melted in a medium frequency induction furnace with a crucible of 30 kg capacity. The vermicularizing treatment was carried out by an Inmold process. The furnace charge consisted of special pig iron containing 0.01% sulphur, and of FeSi75 ferrosilicon, FeMn75 Fe ferromanganese, and technically pure Mo, Cu, Ni and Cr. The cast iron at a temperature of about 1480°C was poured into a sand mould, schematically shown in Figure 1.

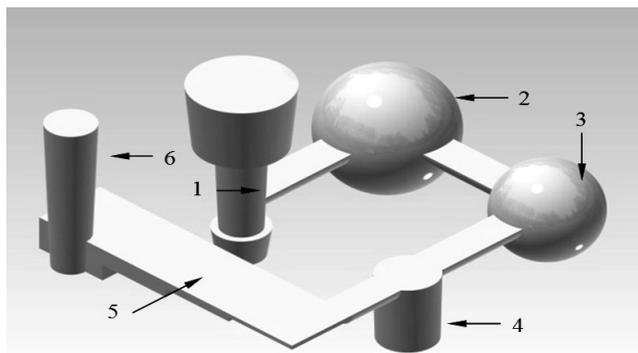


Fig. 1. Schematic layout of mould components for the vermicularizing treatment of cast iron by Inmold process
 1 – downgate, 2 – reaction chamber, 3 – mixing chamber,
 4 – control chamber, 5 – pilot casting, 6 – overflow

In the gating system there was a spherical $\varnothing 85$ mm diameter reaction chamber (2). It contained Lamet® 5504 magnesium master alloy produced by Elkem Company in Norway. The composition of the master alloy is shown in Table 1. The reaction chamber was equipped with a mixing chamber (3) ensuring accurate dissolution of the master alloy and control chamber (4) where the process of the cast iron cooling and solidification was recorded. The trial casting (5) had the shape of steps with 3, 6, 12 and 24 mm wall thickness. The range of chemical composition used in the test cast iron is shown in Table 2. Samples for metallographic tests were cut out from the central part of the stepped casting. The surface fraction of carbides was examined with an NIS-Elements BR image analysis software. Hardness of

cast iron was tested using an HPO-2400 hardness tester operating under the following conditions: ball diameter = $\varnothing 2.5$ mm, load $F = 1840$ N.

Table 1.
 Chemical composition of the master alloy

Chemical composition, wt%					
Si	Mg	Ca	La	Al	Fe
44–48	5–6	0.4–0.6	0.25–0.40	0.8–1.2	rest

Table 2.
 Chemical composition of the CGI tested

Chemical composition, wt%							
C	Si	Mn	Mg	Mo	Cu	Ni	Cr
2.91–	2.28–	0.03–	0.015–	0–	0–	0–	0–
3.82	2.71	1.31	0.023	2.44	3.80	21.04	2,81

3. Data analysis and data mining

Preliminary data analysis was performed on a sample of 180 records. The data concerned different chemical compositions and different wall thicknesses designed for individual compositions. The successive alloying additions were introduced first as single elements and then as mixtures. The calculated correlations have proved the existence of a dependence, clearly indicating at the same time that none of the compounds exhibits a linear trend (Fig. 2.).

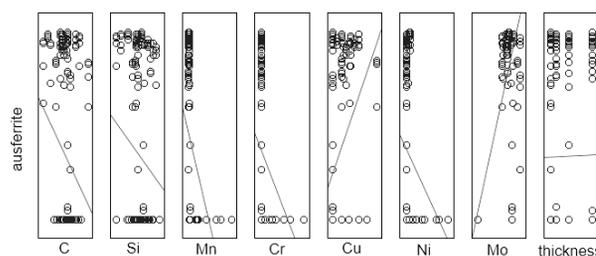


Fig. 2. Matrix form of scatterplots

3.1. Former studies

In previous works, the authors have implemented models of inference based on the obtained results [7]. Forecasting of microstructure based on the chemical composition has been very effective especially for phase constituents such as martensite or austenite (Table 3). Unfortunately, the prediction of ausferrite content has proved to be more difficult.

Table 3.
 Determination coefficient in the prediction models of CGI microstructure

R ²	carbides	pearlite	ferrite	marten site	aus tenite	aus ferrite
ANN	0,96	0,98	0,9	0,99	0,99	0,87
SVM	0,59	0,86	0,7	0,19	0,95	0,81
CART	0,64	0,9	0,71	0,98	0,98	0,83

The basic problem in predicting the content of ausferrite was the selection of a training sample. Experimental data concerned mostly the cases where this microstructural constituent was absent. Predictive models that were designed to forecast the presence of six components were unable to learn and correctly predict the ausferrite content, and it could be said that their accuracy was acceptable only in the situations when ausferrite was not present (Fig. 3).

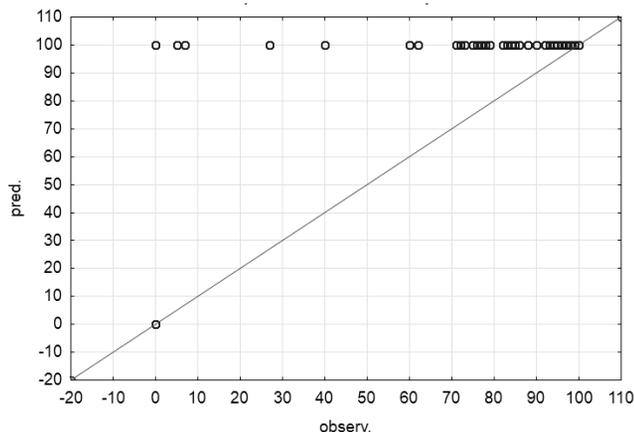


Fig. 3. Incorrect prediction of ausferrite content in models for the six microstructural constituents

On these results was based the decision to develop further prognostic models specialized in the prediction of ausferrite content and apply more adapted data mining techniques to allow learning based on the, so called, weak learners, i.e. explanatory variables that are less correlated with the dependent variable.

3.2. Data mining regression methods

The situation in which a model predicting the value of continuous variable is built is referred to as a regression. Statistical methods for modelling the regression have been developed in recent years fairly intensively. Limits to which is subjected the method of the estimation of linear and nonlinear regression equations are overcome by the use of data mining models that are not subject to assumptions about the density distribution of explanatory variables.

An example of such a technique is **Multivariate Adaptive Regression Splines (MARSplines)** [8,9]. This method does not require assumptions about the distribution of explanatory variables or assumptions about the correlation dependencies between variables, hence results the definition of MARSplines as a non-parametric technique. The dependence is determined by the underlying functions, which in turn are the result of data analysis. The idea of the algorithm is similar to the **Piece Wise Regression (PR)** [10]). The boundaries of the ranges defined for the dependent variable are determined from the data and define the applicability range of particular regression models. MARSplines allows forecasting based on quantitative and qualitative variables [11,12]. The data space is divided into areas where separate regression models and even classification models are implemented. This technique, very popular among the data

mining methods, allows regression based on the multi-dimensional space of the learning vectors [13]. The method is particularly valuable because it offers high precision in situations where dependencies are complex, non-linear, non-monotonic, and thus difficult for parametric modelling. MARSplines is also used successfully in the metal processing industry [14,15].

Artificial neural networks (ANNs), as another technique for solving regression tasks, allow building very precise models [16]. Although ANNs are equally effective in recognizing patterns, and making classification and detection, in the metallurgical industry they are most often used for the mapping processes (metamodelling), which are difficult to describe with numerical models, or the calculation time is too long [17].

ANNs are the mathematical structures of simple computational elements that enable complex transformation of input signals by means of activation functions and multiplication of weights at neuron inputs. They have a multilayered structure, and during training, the error is propagated from the output towards the input to the network, including the optimization of weights at the neuron inputs to match the output signal (network training takes place in iterations called epochs and aims to minimize error function). Once trained network can be used to predict unknown values based on input signals.

The greatest advantage of the networks is their adaptability in situations where there are multiple input signals, the dependencies are complex and unknown, and data may have outliers or deficiencies. Networks are good at generalizing models, but it should be remembered that the results are only approximations.

Neural networks can be used as a benchmark in modelling problems complex and not fully explained. The predictability is high, but no explicit relationships exist between parameters. The network operates on a black box principle- it calculates the results as expected but does not explain how to find a solution. It does not contribute to a better understanding of the phenomenon, either, which is at odds with the goal of data mining research [18].

Random Forest (RF), also known as Bagged Decision Trees) is a method based on the principle of induction of decision trees [19]. It consists in creating complex models composed of multiple decision trees combined into one regression model [20,21]. Trees themselves calculate the value for each successive input, and then the result is averaged. Owing to this approach, the model has finally become independent of outliers, which was previously a major disadvantage of individual decision trees. Random forest is a weak learners team that can solve problems of regression.

The idea of **Boosted Trees (BT)** is also based on a set of decision trees, but the solution is slightly different. In the case of boosted trees, the selected boosted algorithm is used for the construction of subsequent trees. Boosting involves, as in the case of RF, weak learners. At the beginning, the algorithm induces trees without considering the weights of the signals and then increases the weight of observations in the event of a classification error. The next trees are built on the same set of training data, but using the weight of previous trees. This process may require the construction of hundreds of trees in order to minimize the error of an incorrect prediction [22].

Using sets of decision trees can lead to significant improvements in forecasting accuracy.

3.3. Combined techniques

Casting technology still represents a large scale of uncertainty [23], mainly as regards the processes involved in the melt preparation [24,25]. Scientific approaches often refer to methods of reasoning and artificial intelligence [26,27]. These researches implement systems that use formalized knowledge in the form of rules [28]. Frequently, rule-based knowledge is also used in the control devices and software of equipment assisting foundry processes [29-33].

Several regression techniques were used in the study to obtain the greatest possible feedback as regards the knowledge of dependencies. The models used included ANNs; Boosted Trees (BT); Random Forrest (RF); PieceWise Regression (PR); MARSplines (MARS); and also General Linear Models (GLM). All these models have enabled gaining knowledge of the dependencies in a set of variables and knowledge about the influence of individual factors on the ausferrite content.

Initially, models were based on original data from the experiment. Subsequently, a subset of the data was created by selecting only the cases that referred to the ausferrite. This limitation allowed creating models that have learned to predict the specific content of ausferrite and not just confirm its occurrence (Fig. 4).

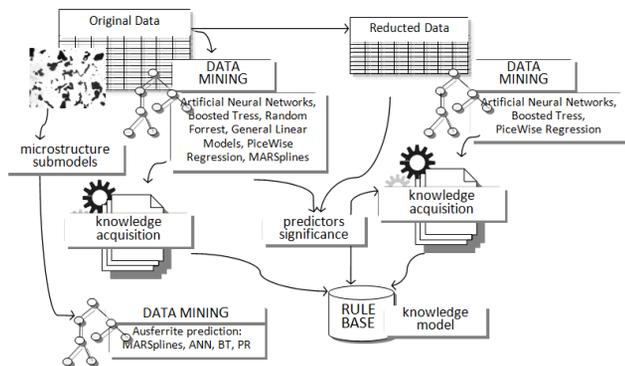


Fig. 4. Data mining process using combined regression techniques

On a reduced sample comprising 72 records, only the models that gave the best results were built. i.e. ANNs; Boosted Trees and PieceWise Regression. Based on the models obtained, knowledge of the dependencies and of the influence of individual alloying constituents was completed, thus creating new rules.

3.4. Results

Data mining techniques have allowed for the development of better or worse prediction models forecasting the ausferrite content. The quality of fit and precision of each model is described in Table 4. However, in addition to the accuracy of the prediction itself, it is important to point out as an important aspect the possibility of obtaining rules (knowledge) about the volatility of individual factors and their influence on the dependent variable.

Table 4.

The quality of fit and precision of each model

	MSE	MAE	r	R ²	RMSE
ANN'180	22,9	2,9	0,99	0,99	4,8
BT'180	30,1	3,4	0,99	0,98	5,5
PR'180	40,0	3,2	0,99	0,98	6,3
GLM'180	48,7	3,3	0,99	0,97	7,0
MARS'180	102,9	6,4	0,97	0,95	10,1
RF'180	130,8	7,2	0,97	0,94	11,4
ANN'72	34,2	4,2	0,97	0,93	5,8
BT'72	56,3	5,3	0,94	0,88	7,5
PR'72	51,8	4,6	0,94	0,89	7,2
ANN'72; 5pr	103,7	7,3	0,87	0,76	10,2
CART'72	349,5	11,0	0,49	0,24	18,7
MARS'72	376,2	13,2	0,42	0,18	19,4
CART submodel ferrite	17,9	1,6	0,68	0,47	4,2
CART submodel pearlite	55,4	3,9	0,79	0,62	7,4
CART submodel carbides	41,3	3,7	0,71	0,50	6,4
CART submodels based	168,7	8,3	0,80	0,64	13,0
MARS submodels based	127,5	8,2	0,85	0,72	11,3
MARS microstructure based	2,3	0,9	0,998	0,995	1,5

It has been reported that ausferrite is formed as a result of molybdenum addition to the alloy, but only in cases where Mo is the alloying additive, the resulting ausferritic structure can occur (Fig. 5). At the same time it has been noted that molybdenum conditions the appearance of ausferrite, but does not determine the level of its content. Manganese, copper and nickel are important in this case (Fig. 6).

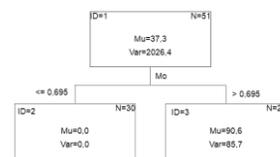


Fig. 5. Tree made with boosted algorithm

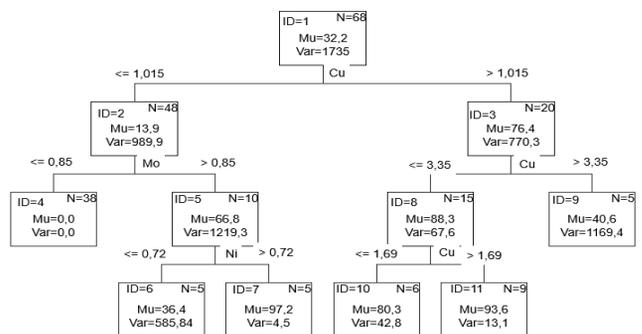


Fig. 6. Tree boosted by weight changes

With the data filtering process, the model precision changes. After removing the relatively easy-to-train (without ausferrite) models, the match coefficients deteriorate, but at the same time considerable improvement is observed in the predictability of actual content levels (Fig. 7).

In the case of boosted trees, a change was noticed as a result of the reduction of data set. On the full set, the algorithm developed more than 800 trees with an error of 30 (Fig. 8), while on the reduced set about 320 trees, but the MSE error almost doubled (Fig. 9).

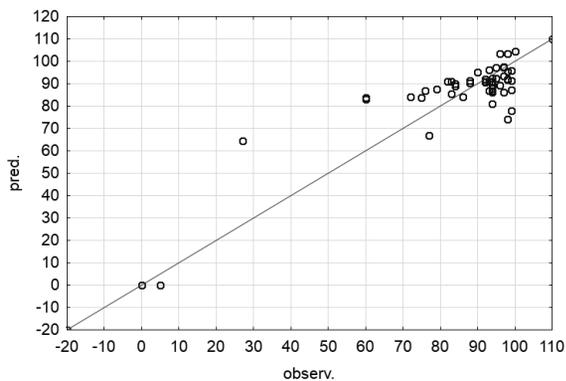


Fig. 7. Measure of PR model prediction fit

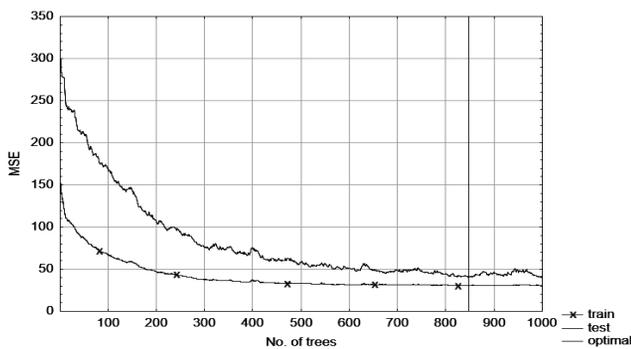


Fig. 8. Cost sequence in BT model for 180 records

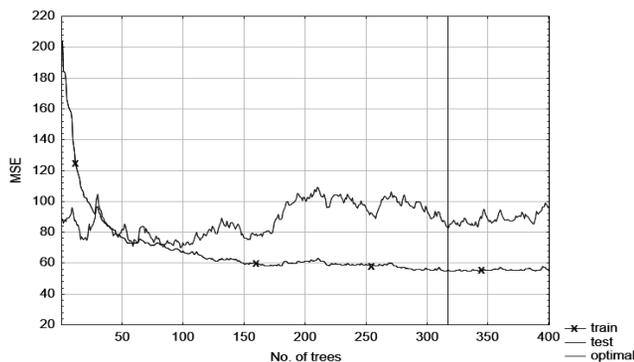


Fig. 9. Cost sequence in BT model for 72 records

During the formation of successive models, the significance of the influence of individual alloying additions and casting wall thicknesses was investigated. Due to the different specifics of models, as well as the different choice of training sets, the results of these calculations differed by showing subtle relationships between variables (Fig. 10).

Studies suggest that ausferrite formation is a two-stage process. The mere presence of various additives is important in the formation of other phases, but it is only the level of their content that is responsible for the transformation promoting the formation of ausferrite. And so, an increase in chromium content favours the formation of carbides; in low concentrations nickel is ferrite- and pearlite-forming element, in the range of 5-9% it favours the formation of martensite, while in higher concentrations it forms austenite. Copper is the addition that most

strongly affects the formation of pearlite. Based on the knowledge about the content of carbides, pearlite and ferrite, it is possible to build a more accurate approximation model. However, since such knowledge is not available *a priori*, a model based on the sub-models (Table 4) has been constructed - it predicts the content of pearlite, ferrite and carbides based on the chemical composition, and next basing on these results, in the second step, it predicts the ausferrite content.

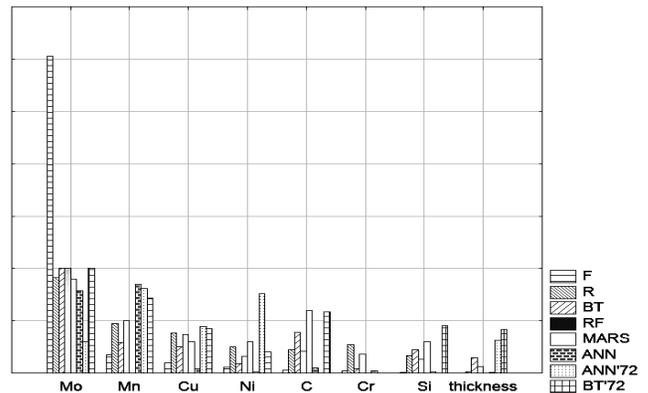


Fig. 10. Standardized effect of variables on the content of ausferrite in individual models

This modelling structure allows mapping of the physical process, but estimation of parameters present in such a model needs to be refined, as the submodel error propagates to the final result of the forecast.

4. Conclusions

The aim of the study was to develop a model of approximation of the content of ausferrite in compacted graphite iron based on the chemical composition and casting wall thickness. As a result of the work, knowledge about the relationship between alloy components and casting microstructure was obtained. The rule base was enriched with conclusions about the strength of the influence of individual alloying elements and also with conclusions about the formation of microstructure and mutual correlations between individual phases. The research used a number of regression modelling methods and also data mining methods, like boosted trees, random forest, neural networks and MARSplines, which helped to judge which of the methods are best in this case of modelling. ANNs offer the highest precision, but models based on decision trees provide the opportunity to gain knowledge in the form of rules.

Acknowledgements

Financial support of The National Centre for Research and Development (LIDER/028/593/L-4/12/NCBR/2013) is gratefully acknowledged.

References

- [1] Skvarenina, S. & Shin, Y.C. (2006). Laser-assisted machining of compacted graphite iron. *International Journal of Machine Tools and Manufacture*. 46(1), 7-17.
- [2] Pietrowski, S. (2000). Compendium of knowledge about compacted cast iron. *Solidification of Metals and Alloys*. 2(44), 279-292. (in Polish).
- [3] Guzik, E. & Kleingartner, T. (2009). A study on the structure and mechanical properties of compacted cast iron with pearlitic-ferritic matrix. *Archives of Foundry Engineering*. 9(3), 55-60.
- [4] Mierzwa, P. & Soński, M.S. (2011). The effect of thermal treatment on the mechanical properties of compacted cast iron. *Archives of Foundry Engineering*. 10(spec.1), 133-138.
- [5] Pytel, A. & Gazda, A. (2014). Evaluation of selected properties in austempered compacted cast iron (AVCI). *Transactions of Foundry Research Institute*. LIV(4), 23-31. DOI: 10.7356/ioid.2014.18.
- [6] Soński, M.S. & Jakubus, A. (2014). Initial Assessment of Abrasive Wear Resistance of Austempered Cast Iron with Compacted Graphite. *Archives of Metallurgy and Materials*. 59(3), 1073-1076. DOI: 10.2478/amm-2014-0183.
- [7] Regulski, K., Wilk-Kołodziejczyk, D. & Gumienny, G. (2016). Comparative analysis of the properties of the Nodular Cast Iron with Carbides and the Austempered Ductile Iron with use of the machine learning and the Support Vector Machine. *The International Journal of Advanced Manufacturing Technology*. 87(1), 1077-1093. DOI: 10.1007/s00170-016-8510-y.
- [8] Friedman, J.H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*. 19(1), 1-67.
- [9] Abraham, A., Steinberg, D. & Philip, N.S. (2001). Rainfall forecasting using soft computing models and multivariate adaptive regression splines. *IEEE SMC Transactions, Special issue on Fusion of Soft Computing and Hard Computing in Industrial Applications*. 1, 1-6.
- [10] Malash, G.F. & El-Khaiary, M.I. (2010). Piecewise Linear Regression: A Statistical Method for the Analysis of Experimental Adsorption Data by the Intraparticle-Diffusion Models. *Chemical Engineering Journal*. 163, 256-263. <http://dx.doi.org/10.1016/j.cej.2010.07.059>.
- [11] Butte, N.F., Wong, W.W., Adolph, A.L., Puyau, M.R., Vohra, F.A. & Zakeri, I.F. (2010). Validation of cross-sectional time series and multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents using doubly labeled water. *The Journal of Nutrition*. 140(8), 1516-1523.
- [12] De Andrés, J., Lorca, P., de Cos Juez, F. & Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: a hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS). *Expert Systems with Applications*. 38(3), 1866-1875.
- [13] Plonsky, L. & Oswald, F.L. (2016). Multiple Regression as a flexible alternative to ANOVA in L2 Research. *Studies in Second Language Acquisition*. 1-14. DOI: 10.1017/S0272263116000231.
- [14] Behera, A.K., Verbert, J., Lauwers, B. & Dufloy, J.R. (2013). Tool path compensation strategies for single point incremental sheet forming using multivariate adaptive regression splines. *Computer-Aided Design*. 45, 3.
- [15] Mukhopadhyay, A. & Iqbal, A., (2009). Prediction of mechanical property of steel strips using multivariate adaptive regression splines. *J App Stat*. 36, 1.
- [16] Jakubski, J., Malinowski, P., Dobosz, St.M. & Major-Gabryś, K. (2013). ANN Modelling for the analysis of the Green Moulding Sands properties. *Archives of Metallurgy and Materials*. 58(3), 961-964.
- [17] Sztangret, L., Szeliga, D., Kusiak, J. & Pietrzyk, M. (2012). Application of inverse analysis with metamodelling for identification of metal flow stress. *Canadian Metallurgical Quarterly*. 51, 440-446.
- [18] Rauch, L., Sztangret, L. & Pietrzyk, M. (2013). Computer system for identification of material models on the basis of plastometric tests. *Archives of Metallurgy and Materials*. 58(3), 737-743. DOI:10.2478/amm-2013-0063.
- [19] Shaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R. & Khovanova (2017). Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. *Biomed. Signal Process Control*. <http://dx.doi.org/10.1016/j.bspc.2017.01.012>.
- [20] Breiman, L. (1999). Random forest. *Mach. Learn.* 45, 1-35, <http://dx.doi.org/10.1023/A:1010933404324>.
- [21] Friedman, J. (2003). Recent advances in predictive (machine) learning. *Proceedings of Physstat*, Stanford University.
- [22] Yang, H., Roe, B.P. & Zhu, J. (2007). Studies of stability and robustness for artificial neural networks and boosted decision trees. *Nucl.Instrum.Meth.* A574, 342-349.
- [23] Kluska-Nawarecka, S., Wilk-Kołodziejczyk, D., Dajda, J., Macura, M. & Regulski, K. (2014). Computer-assisted integration of knowledge in the context of identification of the causes of defects in castings. *Archives of Metallurgy and Materials*. 59(2), 743-746.
- [24] Mrzygłód, B., Kowalski, A., Olejarczyk-Woźńska, I., Adrian, H., Głowacki, M. & Opaliński, A. (2015). Effect of heat treatment parameters on the formation of ADI microstructure with additions of Ni, Cu, Mo. *Archives of Metallurgy and Materials*. 60(3A), 1941-1948.
- [25] Olejarczyk-Woźńska, I., Adrian, H. & Mrzygłód, B. (2014). Mathematical model of the process of pearlite austenitization. *Archives of Metallurgy and Materials*. 59(3), 981-986.
- [26] Kluska-Nawarecka, S., Gorny, Z. & Wilk-Kołodziejczyk, D. (2007). The logic of plausible reasoning in the diagnosis of castings defects. *Archives of Metallurgy and Materials*. 52(3), 375-380.
- [27] Warmuzek, M. & Regulski, K. (2011). A procedure of in situ identification of the intermetallic AlTMSi phase precipitates in the microstructure of the aluminum alloys. *Practical Metallography*. 48(12), 660-683.
- [28] Kluska-Nawarecka, S., Wilk-Kołodziejczyk, D., Regulski, K. & Dobrowolski, G. (2011). *Rough sets applied to the RoughCast system for steel castings*, Intelligent Information and Database Systems. Part II, Nguyen, N.T., Kim, Ch.G., Janiak, A. (Eds.), Springer Lecture Notes in Computer Science 6592, 52-61.
- [29] David, J., Švec, P., Garzinová, R., Kluska-Nawarecka, S., Wilk-Kołodziejczyk, D. & Regulski, K. (2016). Heuristic modeling of casting processes under the conditions uncertainty. *Archives of Civil and Mechanical Engineering*. 16(2), 179-185.
- [30] David, J., Jancikova, Z., Frischer, R. & Vrožina, M. (2013). Crystallizer's Desks Surface Diagnostics with Usage of Robotic System. *Archives of Metallurgy and Materials*. 58(3), 907-910.
- [31] Smyksy, K., Ziółkowski, E., Wrona, R. & Brzeziński, M. (2013). Performance evaluation of rotary mixers through monitoring of power energy parameters. *Archives of Metallurgy and Materials*. 58(3), 911-914.
- [32] Glowacz, A. & Glowacz, Z. (2017). Recognition of rotor damages in a DC motor using acoustic signals. *Bulletin of the Polish Academy of Sciences Technical Sciences*. 65(2), 187-194.
- [33] Maciol, P. Gawad, J. & Podorska, D. (2007). Arrangement of flow modification devices in continuous casting tundish based on multicriterion optimization. *Archives of Metallurgy and Materials*. 52 (1), 105-112.