DE GRUYTER OPEN

Arch. Min. Sci., Vol. 59 (2014), No 3, p. 705–720

Electronic version (in color) of this paper is available: http://mining.archives.pl

DOI 10.2478/amsc-2014-0049

JACEK JAKUBOWSKI*, ANTONI TAJDUŚ*

PREDICTIVE REGRESSION MODELS OF MONTHLY SEISMIC ENERGY EMISSIONS INDUCED BY LONGWALL MINING

REGRESYJNE MODELE PREDYKCYJNE MIESIĘCZNEJ EMISJI ENERGII SEJSMICZNEJ INDUKOWANEJ EKSPLOATACJĄ W ŚCIANIE

This article presents the development and validation of predictive regression models of longwall mining-induced seismicity, based on observations in 63 longwalls, in 12 seams, in the Bielszowice colliery in the Upper Silesian Coal Basin, which took place between 1992 and 2012. A predicted variable is the logarithm of the monthly sum of seismic energy induced in a longwall area. The set of predictors include seven quantitative and qualitative variables describing some mining and geological conditions and earlier seismicity in longwalls. Two machine learning methods have been used to develop the models: boosted regression trees and neural networks. Two types of model validation have been applied: on a random validation sample and on a time-based validation sample.

The set of a few selected variables enabled nonlinear regression models to be built which gave relatively small prediction errors, taking the complex and strongly stochastic nature of the phenomenon into account. The article presents both the models of periodic forecasting for the following month as well as long-term forecasting.

Keywords: Induced seismicity, mining tremors, rockburst hazard, longwall mining, boosted trees, neural networks, data mining, regression models, predictive models

W artykule przedstawiono budowę i walidację predykcyjnych modeli regresyjnych sejsmiczności indukowanej eksploatacją w ścianie, opartych na obserwacjach w 63 ścianach kopalni Bielszowice prowadzonych w 12 pokładach w latach 1992-2012. Zmienna prognozowaną jest logarytm miesięcznej sumy energii sejsmicznej wstrząsów w ścianie. Zestaw predyktorów składa się z siedmiu zmiennych ilościowych i jakościowych opisujących wybrane czynniki górnicze i geologiczne w ścianach. Do budowy modeli zastosowano dwie metody uczenia się maszyn: drzewa wzmacniane oraz sieci neuronowe. Zastosowano dwa rodzaje walidacji modeli: na losowej próbie walidacyjnej oraz na czasowej próbie walidacyjnej.

Zestaw kilku wybranych zmiennych pozwolił na zbudowanie nieliniowych modeli regresyjnych, które, biorąc pod uwagę złożoną i silnie stochastyczną naturę zjawiska, dają względnie małe błędy pro-

* AGH UNIVERSITY OF SCIENCE & TECHNOLOGY, DEPARTMENT OF GEOMECHANICS, CIVIL ENGINEERING & GEOTECHNICS; AL. A. MICKIEWICZA 30, 30-059 KRAKOW, POLAND

gnozy. W artykule przedstawiono zarówno modele do prognozy okresowej na kolejny miesiąc jak i do prognozy długoterminowej.

Słowa kluczowe: Sejsmiczność indukowana, wstrząsy górnicze, zagrożenie tąpaniami, eksploatacja ścianowa, drzewa wzmacniane, sieci neuronowe, data mining, modele regresyjne, modele predykcyjne

1. Introduction

Rock mass at great depth is found in a state of primary, stable mechanical equilibrium, shaped by high vertical pressure, by geological history and by local tectonic structures. Mining activity disturbs the primary mechanical equilibrium of rock mass, inducing movement, stress and elastic energy concentrations. Local ruptures and sudden discharges of elastic energy take place along with propagation of seismic waves, in other words – tremors, and sometimes rockbursts accompanied by excavation destruction.

The most important and most difficult geophysical and mining research problems include forecasting the size, time of occurrence and location of these tremors, as well as the level of seismic activity in a given area and during a given time interval.

A few research tools are used to build short- and long-term seismicity forecasts such as the Gutenberg-Richter relation and the variations of its parameters (Lasocki, 1993; Gołda & Kornowski, 2011), probability distributions (Lasocki, 1992; Idziak et al., 1991), nonparametric estimators of probability distributions (Lasocki & Orlecka-Sikora, 2008; Orlecka-Sikora, 2008), fractal description of energy, time, and the locations of tremor centres and the variation of the fractal dimension (Idziak & Zuberek, 1995), space-time cluster analysis (Leśniak & Isakow, 2009), relationship between seismic energy and extracted output (Głowacka, 1993; Stec, 2008; Jakubowski, 2014), pattern recognition (Marcak, 1993), neural networks (Kabiesz, 2008; Jakubowski, 2014), logistic regression and boosted trees (Jakubowski, 2014), rule induction methods (Sikora, 2011), time series analysis considering the seismic and acoustic energy (Kornowski & Kurzeja, 2008; Cianciara & Cianciara, 2005, 2006).

This article presents a regression predictive model of monthly seismic energy induced in a longwall area, based on two machine learning methods: gradient boosted regression trees and neural networks. This is a nonparametric data-based model and not a physical model. The forecast is a result of the development and implementation of a model which describes the relationship between 7 continuous and categorical predictors and the continuous predicted variable.

2. Geological and mining conditions and seismic activity in Bielszowice colliery

The Bielszowice colliery, which lies in the southern part of Poland, in the north-western part of the Upper Silesian Coal Basin, within the boundaries of the cities of Zabrze, Ruda Śląska and Mikołów is comprised of 34 km^2 mining area. The mine employs 3.5 thousand people and has an output of about 2 million tonnes of coal a year.

The deposit is intensively cut with faults with latitudinal, longitudinal and diagonal orientation. A number of faults have significant thrust (up to 200 m) and create local faulting zones (up to 220 m wide). There are also numerous local faults with thrusts of several metres and some other tectonic features and seam irregularities. The main faults with latitudinal orientation divide the deposit into three parts: northern, middle (so-called Centralna part where seams from 410 to 510 are mined) and southern (so-called Borowa part, has available seams which range from 358/1 to 408). The main drifts and respective protection pillars run along the longitudinal axis of the mining area and divide it into the eastern and western parts. The average strata inclination is 10° and locally even 17°.

The Bielszowice mine is among mines associated with high rockburst hazards, with 27 rockbursts recorded in the last 35 years. The majority of the series 500 and 400 seams are classified in the highest rockburst hazard category. Owing to this hazard, the Bielszowice mine utilizes a stress-relieving extraction system across a long, regular front. In addition to the rockburst hazard, there are also methane and fire hazards as well as difficult climatic conditions.

All of the 63 longwalls considered, were mined between 1992-2012 with roof caving in the gob, in 12 seams of the Rudzkie layers (400 series) and the Siodłowe layers a (500 series), these being in the Westphalian and Namurian stages of the Carboniferous period.

Figures 1-4 present the quantitative characteristics of the extraction in longwalls based on monthly-aggregated data. In particular, there are histograms for monthly longwall output,

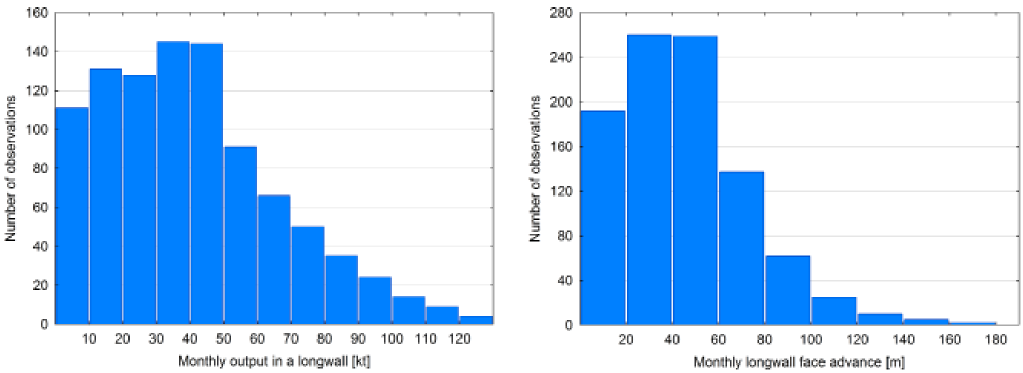


Fig. 1. Monthly output histogram (left) and monthly advance in longwalls histogram (right)

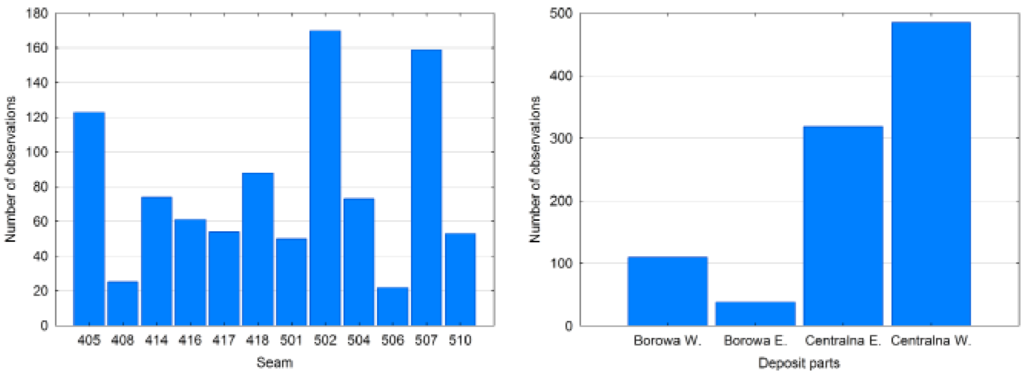


Fig. 2. Number of observations by seams (left) and by parts of mine deposit (right)

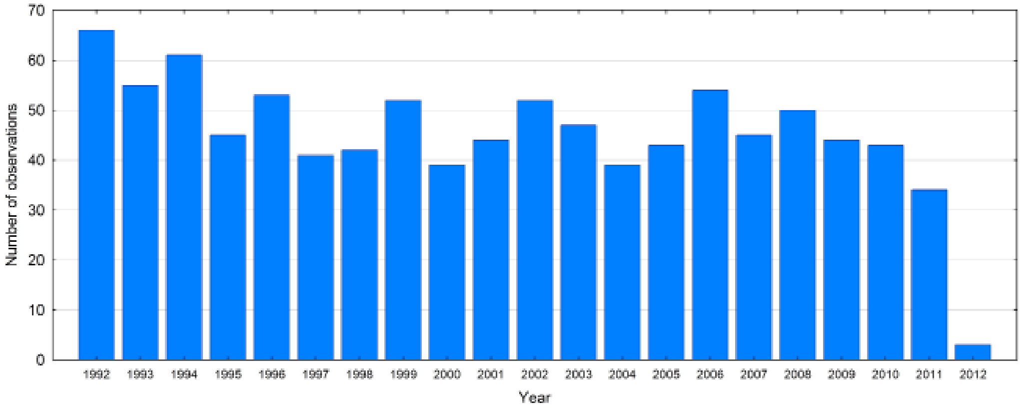


Fig. 3. Number of observations by year

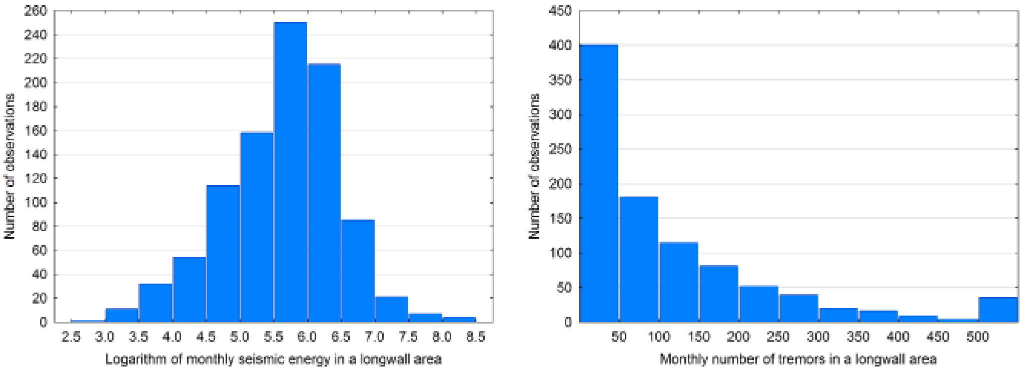


Fig. 4. Logarithm of monthly seismic energy histogram (left) and monthly number of tremors histogram (right)

advance, number of tremors and seismic energy. The number of extraction months in layers, sections and years are also presented. One observation on the histogram indicates one month of extraction in one longwall.

3. General approach and analytical methods applied to the model development

Classic statistical inference uses mathematical tools for the precise evaluation of uncertainty. Data and models must comply with rigorous assumptions and requirements. In the data mining approach, uncertainty evaluation is not strict, but empirical. Due to this, generalizations to the entire population are not strict and may be evaluated quantitatively, but only empirically, by means of verification.

The empirical uncertainty evaluation in machine learning methods is based on the division of the dataset into training, testing, and validation sets. The model is built on the training set with the involvement of the testing set and evaluated on the validation set (Fig. 5).

In the classical statistical approach, the general form of the relationship between the variables under study is usually assumed in advance. Development of the model is the estimation of its interpretable parameters. This is not the case in the majority of machine learning models. They form the black box models constructed solely on the basis of data, the models whose parameters lack a direct, physical interpretation.

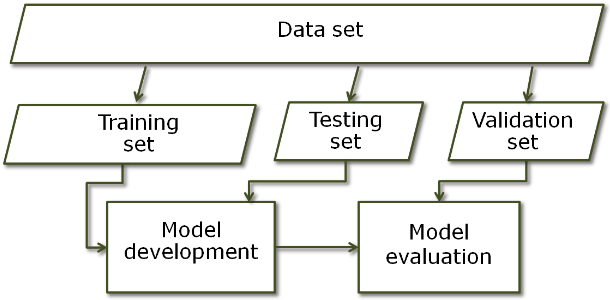


Fig. 5. Model development and evaluation in data mining and machine learning methods

Two machine learning methods have been used here to develop models: boosted regression trees and neural networks. The final forecast is an arithmetic mean of the forecasts obtained using each of the two methods. Both component models were built using the specialized Statistica Data Miner v.10 (StatSoft, 2011) system. Extensive descriptions of algorithms used in these methods, interpretation of parameters, advantages and disadvantages can be found in publications (Berry & Linoff, 2000; Seidman, 2000; Hand & Mannila, 2001; Larose, 2008; Lasek, 2002; Ratner, 2003; Koronacki & Ćwik, 2005; Tadeusiewicz et al., 2007; Migut & Haranczyk, 2011; StatSoft, 2011). Below, only an outline of general model development rules in both methods is presented.

Boosted regression trees model consist of many simple C&RT trees. The component trees entered into the model are generated iteratively. The first tree was created using training set data, while the next was created on the basis of training data modified with weights in such a way as to represent erroneously classified cases with greater weights. In each consecutive step, case weights are corrected and a new component tree is generated. The response of the trained model is the averaged response of all component trees. The process of adding new component trees to the model is only interrupted when the model error on the test sample ceases to decrease.

The second analytical method to be applied was a neural network – a feed forward neural network with linear aggregating function, a so-called multilayer perceptron. Every neuron on each layer is connected to every neuron on the neighbouring layers, with a weight corresponding to each connection. The purpose of training the network is to find a weight vector for which the network's output values are the closest to those observed, or the measure of error is the smallest. The process of training a neural network consists of many successive stages (epochs), during which, an entire training set is scanned, the weight vector is then modified and thus error is minimized. The learning error is minimized on the learning sample and monitored on the test sample in order to prevent the network over-learning. Initial weight values are chosen at random (StatSoft, 2011).

Table 1 presents the basic parameters of the models developed by means of both aforementioned techniques, using models described in section 5 as examples.

TABLE 1

Basic parameters of boosted regression trees and neural networks models. Specifications refer to models described in section 5 as examples

Models	Model specifications	
Neural networks	Number of input neurons	8
	Number of hidden neurons	10
	Number of output neurons	1
	Number of epochs	144
	Training algorithm	Broyden, Fletcher, Goldfarb, Shanno
	Error function	Sum of squares
	Hidden activation	Hyperbolic tangent
	Output activation	Hyperbolic tangent
Boosted regression trees	Optimal number of trees	199
	Maximum tree size	5
	Learning rate	0.1
	Random test data proportion	0.3
	Subsample proportion	0.5

4. Input data, predicted variable and predictors

Input data include data from the seismic catalogue, basic data on the longwall extraction, and characteristics of mining and geological conditions accompanying the exploitation.

Data from seismic catalogue include about 115,000 tremors recorded in the vicinity of 36 longwalls between 1992 and 2012. Longwall extraction data include, inter alia, monthly output, basic longwall parameters such as face length, height, depth, seam and region number, mining scheme etc. Other data describes mining and geological conditions in longwall face location and in the overlying and underlying seams.

Data on tremors and mining and geological conditions were collected in data files with different structures, so they had to be merged and aggregated by month separately for each longwall. Months with no mining or no tremors were disregarded. The pre-processed data included nearly 1000 cases and over 200 continuous and categorical variables initially considered in the analysis.

The dependent (predicted) variable is a logarithm of the monthly sum of seismic tremor energy in a longwall. Seven dependent variables, including 6 quantitative and 1 binary, were selected by examination and testing of many variable sets and models.

The independent variables (predictors) in the models described below comprise:

Var1: Total longwall face advance [m],

Var2: Uniaxial compressive strength of coal in the seam [MPa],

Var3: Total MRG method score for a longwall,

Var4: Longwall face distance to the nearest fault with thrust exceeding 3 m [m],

Var5: Impact of extraction edges for current longwall location,
Var6: Longwall advancing along faults (yes/no, binary variable),
Var7: Logarithm of monthly seismic energy in the longwall in previous month.

Total longwall face advance is the distance travelled by the face from the set-up position to the current one. MRG (in Polish "Metoda Rozeznania Górniczego") is an expert method of rocburst hazard evaluation, comprised of several simple geological, geomechanical and mining technology feature scores. It is commonly used as a supporting element of the comprehensive rockburst hazard evaluation in Polish coal mining (Barański et al., 2007). 'Impact of extraction edges' refers to gob-solid edges, pillars and remnants in the overlying and underlying coal seams. It is calculated based on previous mining activities documented in neighbouring seams and depends both on vertical distance and time of extraction. If the considered longwall pannel was adjacent and advancing parallel to fault or fault zone, the binary variable *Var6* equals 1.

Figures 6 through to 8 present histograms of independent variables in the dataset. One observation on the histogram indicates one month of extraction on one longwall.

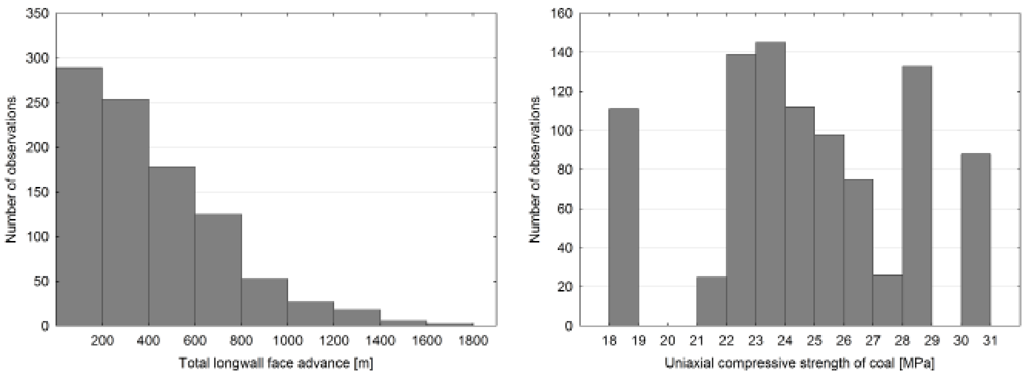


Fig. 6. Total longwall advance histogram (left) and uniaxial compressive strength of coal in seams histogram (right)

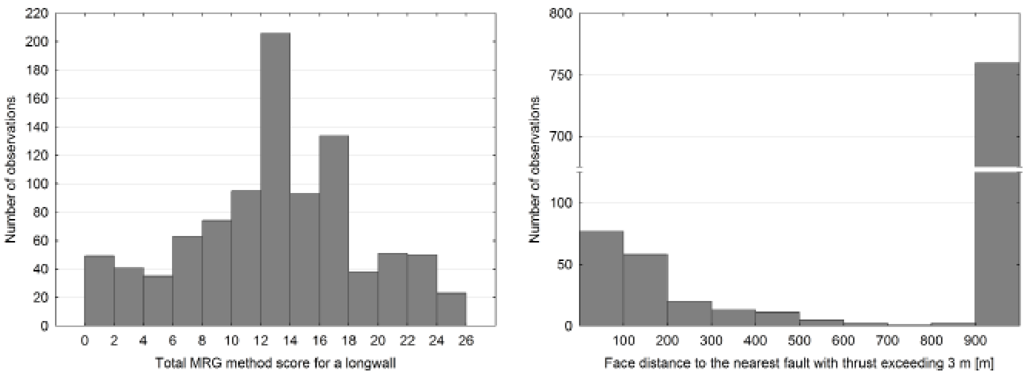
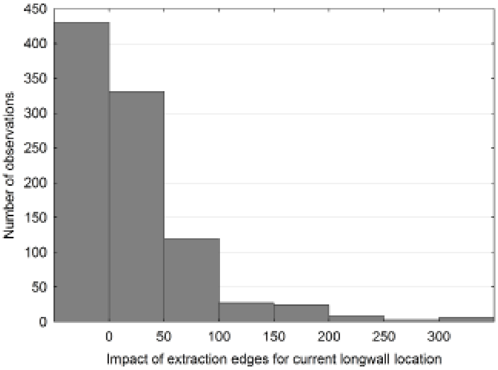


Fig. 7. Total MRG method score histogram (left) and distance to the nearest fault with thrust exceeding 3 m histogram (right)



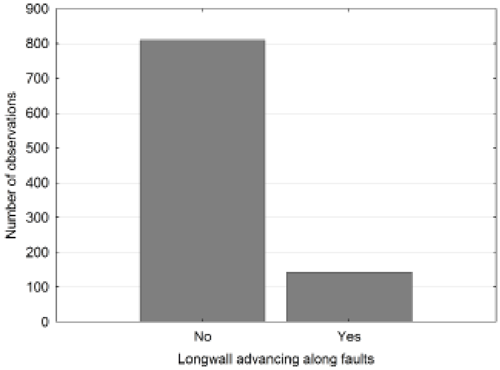


Fig. 8. Impact of extraction edges histogram (right) and histogram of binary variable informing whether the longwall was advancing along faults (left)

Table 2 shows the correlation matrix of variables. The relationships between variables are strongly nonlinear and the linear correlation analysis assumptions are mainly not met, but such a table could still be expected and helpful at the stage of preliminary data analysis.

TABLE 2

Correlation matrix for dependent and continuous independent variables. *Target* denotes the dependent variable. Independent variables are designated *Var1-Var7* as in the description above. *Var6* is not included because it is a categorical variable.

	Target	**Var1**	**Var2**	**Var3**	**Var4**	**Var5**	**Var7**
Target	1.00	–0.07	0.25	0.31	–0.22	–0.01	0.50
Var1	–0.07	1.00	–0.13	–0.09	0.01	0.01	0.25
Var2	0.25	–0.13	1.00	–0.36	–0.23	–0.10	0.15
Var3	0.31	–0.09	–0.36	1.00	–0.28	0.17	0.18
Var4	–0.22	0.01	–0.23	–0.28	1.00	0.06	–0.19
Var5	–0.01	0.01	–0.10	0.17	0.06	1.00	0.03
Var7	0.50	0.25	0.15	0.18	–0.19	0.03	1.00

Six independent variables include data determined on the basis of extraction plans and recognized faults, hence the data are available in advance in relation to mining. A single independent variable (logarithm of the monthly seismic energy in previous month) can only be determined during the mining extraction, at the end of each month, and used in forecasts for the next month.

Using the longwall output as one of the predictors did not significantly improve the monthly seismic energy model, contrary to the daily seismic energy models investigated elsewhere (Jakubowski, 2014).

The characteristics of the problem, along with the dataset size are suitable for development and evaluation of regression models, using each of the applied analytical methods.

5. Development and evaluation of the regression model validated on a random sample (A)

The aforementioned machine learning methods were both used to develop the model. The methods were used in the data mining approach, i.e. according to the concept presented in Figure 5. It was assumed that the final forecast was an arithmetic mean of predictions received from component models.

The development of these models was preceded by a preliminary data review, data cleaning, missing data handling and defining dependent variable according to the established problem. The choice of predictors was performed during the process of repeated analysis and evaluation of models.

The dataset was divided into training and testing sets on which the models were built, and an independent, random validation set on which the models were verified. Table 3 presents goodness of fit for component models and the final model. Figure 9 includes prediction-observation scatterplots for these models. Measurements of goodness of fit in the validation set are close to those in the training-testing set, thus indicating the stability of the final model. High R 0.80 and R^2 0.64 values indicate that the model describes a significant portion of variability of monthly seismic energy. Taking into account the complex and strongly stochastic nature of the phenomenon, the limited range of data and the small number of predictors, the evaluation of the model is high.

A certain narrowing of prediction range in relation to the observation range can be noticed for low seismic energy values.

Various measures of predictor importance have been developed for different machine learning methods. Figure 10 shows some measures of importance of predictors used for the boosted trees and neural network methods (StatSoft 2011) which are useful during the development of the models. The higher the values on the vertical axes, the 'more important' the variable is for the model. Note, that according to the measures presented, *Var6* is the most important variable in the neural networks model and the least important variable in the regression boosted trees model (Fig. 10). *Var1* is the second most important variable in the boosted trees model and the second least important variable in the neural networks model. Prediction quality is comparable with both models. Both analytical methods applied are black-box methods, hence the direct, physical interpretation of these graphs may be misleading.

The magnitude of estimated prediction errors indicates that the model can be used in practice to predict the monthly sum of seismic energy in a longwall. Please note that the same model can be used to predict the monthly seismic energy in longwalls located in different seams and mine sections.

TABLE 3

Criteria and results of evaluation of component models and final model A

	Training and testing sets, 1992-2012				Random validation set, 1992-2012			
Models	**Size**	**Correlation coefficient**	**Mean squared error**	**Mean absolute error**	**Size**	**Correlation coefficient**	**Mean squared error**	**Mean absolute error**
Regression boosted trees	721	0.83	0.22	0.35	231	0.77	0.32	0.42
Neural networks	721	0.82	0.23	0.35	231	0.80	0.29	0.39
Final model A	**721**	**0.84**	**0.21**	**0.34**	**231**	**0.80**	**0.29**	**0.39**

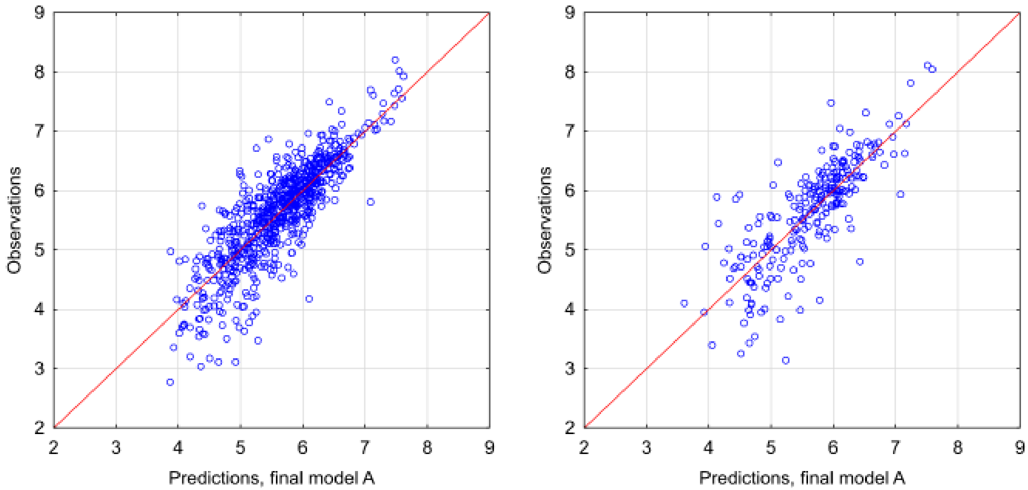


Fig. 9. Observation-prediction scatterplot for the final model A on the training-testing set (left) and random validation set (right)

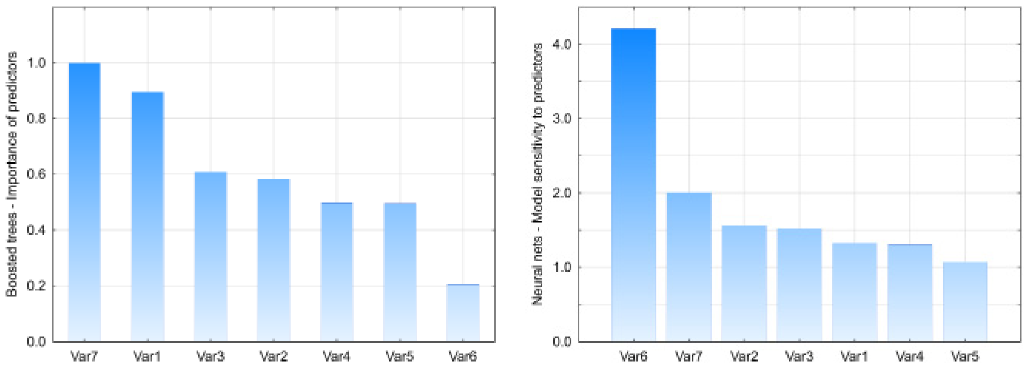


Fig. 10. Predictor importance measures for component models: boosted regression trees (left) and neural networks (right)

6. Regression models with validation on time-based sample (B1 and B2)

The independent validation set, against which the models were evaluated in the preceding section, was uniformly sampled from a dataset and included cases from various seams, sections and longwalls, covering the entire data period. This therefore allowed for the proper verification and evaluation of model quality. It allowed for the presumption that the model would also be able to reflect conditions which could occur in successive months of mining.

However the sample, comprising successive months of mining, is not a random sample from the whole dataset, but a time-based sample covering a certain period in the future.

One might therefore be concerned that, as a result of changes in mining and geological conditions, which may have taken place during the analysed twenty year period of exploitation, models trained on the whole dataset may be irrelevant to a certain time interval in the future.

Therefore, the models' behaviour was checked against the time-based validation sample. For this purpose, data from the final period of exploitation, between 2010 and 2012 (lasting more than two years) was separated to serve as a time-based validation sample. The remaining data from 1992-2009 were used to develop model B1. As with the previous case, the final prediction is an average of the two component models. Evaluation of the final model B1 is shown in table 4 and Fig. 11.

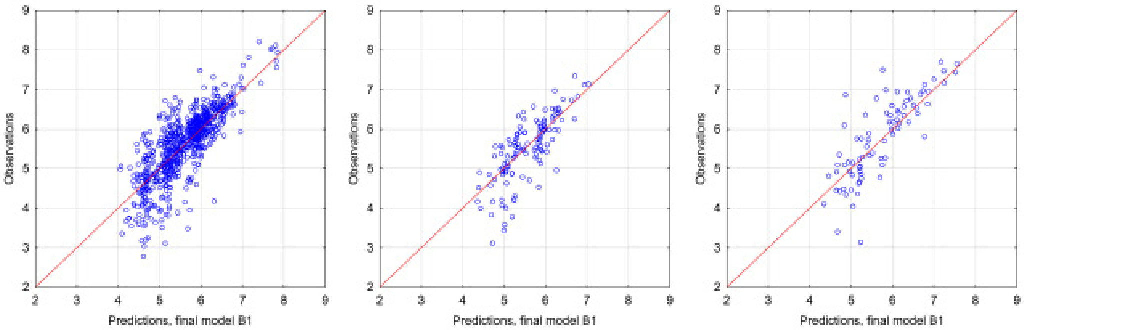


Fig. 11. Observation-prediction scatterplots for model B1. Results on the training-testing sample 1992-2009 (left), random validation sample (middle) and time-based validation sample 2010-2012 (right)

In addition, a similar procedure was performed for the time-based validation set corresponding to the initial exploitation period between 1992 and 1993. The remaining data from 1994-2012 were used to develop model B2. Evaluation of the final model B2 can be found in table 4 and Fig. 12.

TABLE 4

Criteria and results of evaluation of models B1 and B2 on time-based validation set

	Training and testing sets					**Validation sets**				
Models	**Model built using the data from**	**Size**	**Correlation coefficient**	**Mean squared error**	**Mean absolute error**	**Validation type and data range**	**Size**	**Correlation coefficient**	**Mean squared error**	**Mean absolute error**
Final model B1	1992-2009	749	0.79	0.26	0.37	Random 1992-2009	123	0.76	0.29	0.39
						time-based 2010-2012	80	0.81	0.37	0.44
Final model B2	1994-2012	716	0.80	0.28	0.38	random 1994-2012	115	0.80	0.27	0.36
						time-based 1992-1993	121	0.81	0.18	0.32

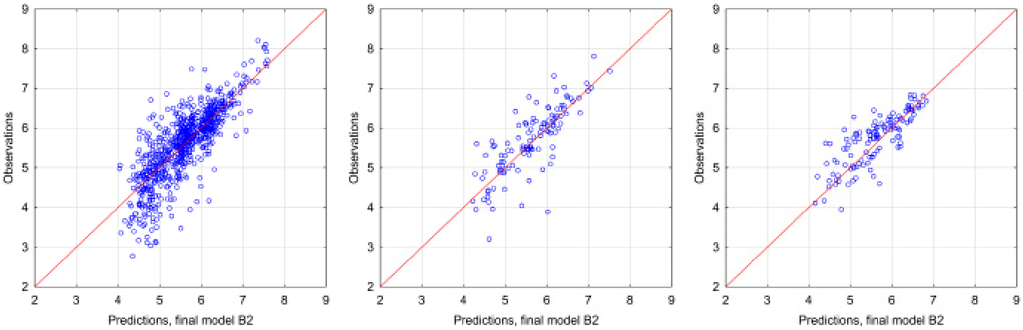


Fig. 12. Observation-prediction scatterplots for model B2. Results on the training-testing sample 1994-2012 (left), random validation sample (middle) and time-based validation sample 1992-1993 (right)

It turns out that both models (B1 and B2) work well on time-based validation sets covering the longwall extraction in the two outermost time intervals.

The observation-prediction correlation coefficients on both time-based validation sets are high and comparable to the values obtained with random validation sets and training-testing sets (Table 4). Error values on the respective sets are also similar and moderate in general. The correlation coefficients and model errors found in time-based validation sets do not differ much from the results obtained from model A which was developed and verified on data from 1992-2012. Hence, it may be inferred that for the conditions prevailing in the Bielszowice mine, the models trained on data from 18 years of exploitation are stable in the following 2-year period and that the description of phenomenon by means of the variables selected for the model is capable of reflecting changes in mining and geological conditions occurring in that perspective.

As shown by training the model on the data from 1992-2009 and verifying it using the data from 2010-2012, and by developing the model on the data from 1994-2012 and verifying it using the data from 1992-1993, the model can be trained using available historical data and used for future predictions.

7. Regression model for a long-term forecast (C)

Earlier regression models are based, inter alia, on tremor data from previous months. Such models can be used, for instance, during the next 2-year period (or updated every three months or every year) as a predictive model for periodic, monthly forecasts. This means that a forecast for the next month can be made at the end of each month after summarizing tremor energy data.

The question arises as to whether it's possible to build a good predictive model based solely on mining and geological characteristics without taking into account the seismic energy in the previous month. Such a model could be used to predict the monthly seismic energy emissions several months or a couple of years in advance, which would be an advantage.

Similarly to the previous models, a regression model was developed using the same set of 6 variables and without the *Var7* variable describing the logarithm of the seismic energy from the previous month. The model evaluation on the validation set is presented in Table 5 and Figure 13.

TABLE 5

Criteria and results of evaluation of component models and final model C

	Training and testing sets, 1992-2012				Random validation set, 1992-2012			
Models	**Size**	**Correlation coefficient**	**Mean squared error**	**Mean absolute error**	**Size**	**Correlation coefficient**	**Mean squared error**	**Mean absolute error**
Regression boosted trees	809	0.71	0.34	0.46	143	0.67	0.52	0.55
Neural networks	809	0.73	0.32	0.44	143	0.74	0.41	0.50
Final model C	809	0.75	0.31	0.43	143	0.74	0.43	0.50

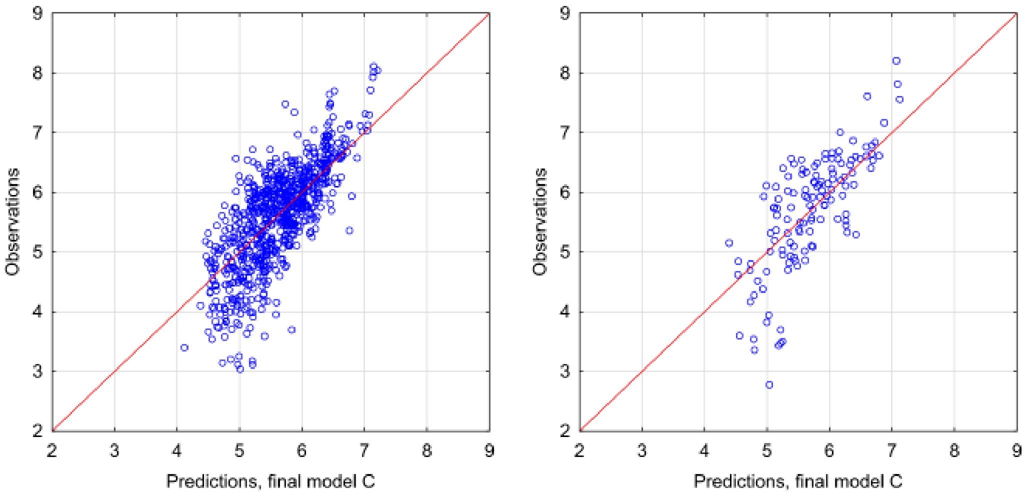


Fig. 13. Observation-prediction scatterplots for the final average model C. Results on the training – testing sample (left) and validation sample (right)

It turns out that model C works rather well. Correlation coefficients are lower than the correlation coefficients for model A. Its mean squared error and mean absolute error are moderately greater than the errors in model A. Taking a long, possibly 2-3 year perspective for the model into account, the forecast errors seem to be reasonable.

Some narrowing of the prediction range in relation to the observation range can be detected, which results in deflated high values and inflated low values of energy predictions (Fig. 13).

Predictor *Var7* was indicated as important for both component models of the final model A (Fig. 10) and it had the greatest linear correlation coefficient with the dependent variable (Tab. 2). The lack of this predictor in model C increased prediction error, but much less than it could be expected.

8. Summary

This article presents the development and validation of predictive, nonlinear, nonparametric regression models of longwall mining-induced seismicity, developed on the basis of data for mining carried out for 63 longwalls in the Bielszowice colliery between 1992 and 2012.

The predicted variable was the logarithm of the monthly sum of seismic tremor energy in the longwall area, and the predictors were the variables which characterize the mining and geological conditions for the longwalls and the induced seismicity. Two regression methods have been applied: regression boosted trees and neural networks. Each final model is an average of the component models obtained using these two analytical methods.

The first model (A), presented in section 5, was for periodic, monthly forecasts. Low estimated prediction errors indicate that the model may be useful in forecasting monthly seismic energy in the longwall. Quite high correlation and determination coefficients on the validation set indicate that the model describes a significant portion of variation in monthly seismic energy. The same model can be used to predict monthly seismic energy in longwall panels located in various seams and parts of the Bielszowice colliery deposit.

Any forecast, using any method, is based on the assumption that general dependencies on which the model is based will be valid during the forecast period. A decision was made to check whether this assumption was true in the 1992-2012 period. The regression model (B1) was developed using data from 1992-2009 and it was used to make a forecast for 2010-2012. In other words, it was checked whether the model developed at the end of 2009 could be used to predict seismic energy in the next 2 years, or if a possible change in mining and geological conditions in 2010-2012 would render the model irrelevant.

In addition, a similar verification was performed "backwards", i.e. it was checked how model (B2) developed using the data from 1994-2012 estimated the seismic energy from 1992-1993.

It turned out that both models (B1 and B2) work well and are stable for the data coming from 2-year validation periods. Hence, it can be presumed that the phenomenon description using the variables chosen for the model and the time interval of training data make it possible to account for significant factors which affect monthly seismic activity and to explain a portion of observed variability of monthly seismic activity.

It was therefore demonstrated that good and stable regression models can be developed for the Bielszowice mine and that such models can be used during the next 2 years of mining for periodic forecasts of monthly seismic energy in various longwalls for the following month.

It was then decided to check whether it is possible to make a reasonably good forecast not just one month in advance, but several months or a couple of years in advance. The model used for such a prediction had to be based only on data available at the time the forecast was made, so it could not include the seismic energy in the previous month.

The model for a long-term forecast (C) is inferior to the previous models, but still works quite well. It has lower correlation coefficients than model A. Its mean squared error and mean absolute error are moderately greater than the errors in model A. Taking into account a long perspective for the model, the forecast errors seem to be acceptable.

The set of several chosen independent variables describing mining conditions in the Bielszowice mine made it possible to develop reasonably good regression models for predicting the logarithm of the total monthly seismic energy. Taking the complex and strongly stochastic nature of the phenomenon into account, the models have relatively small prediction errors and high correlation coefficients. A certain narrowing of the prediction range in relation to the observation range occurs in various models to a different degree.

The monthly induced seismic energy describes the longwall propensity for seismic emissions, but is not a direct measure of the hazard of extremely large tremors or rockburst hazard. The models presented in this article are not designed for the prediction of these phenomena. They can be used to forecast monthly mining-induced seismic energy in the Bielszowice mine, both as a periodic forecast for the following month or a long-term forecast. The comprehensive evaluation of the approach presented as well as its utility, requires further research, however the results presented for Bielszowice mine are encouraging.

Acknowledgements

The authors wish to thank the management and employees of Kompania Węglowa and Bielszowice colliery for their support in collecting data necessary for the research.
The authors are grateful to Michał Zelek and Łukasz Wanat for collecting data used for this work.
Prepared within research 11.11.100.197

References

Barański A., Drzewiecki J., Kabiesz J., Konopko W., Kornowski J., Krzyżowski A., Mutke G., 2007. *Zasady stosowania metody kompleksowej i metod szczegółowych oceny stanu zagrożenia tąpaniami w kopalniach wegla kamiennego*. GIG Seria Instrukcje, No 20, Katowice.

Berry M., Linoff G., 2000. *Mastering Data Mining*. Wiley, Hoboken, NJ.

Cabena P., Hadjinian P., Stadler R., Verhees J., Zanasi A. 1998. *Discovering Data Mining: From Concept to Implementation*. Prentice Hall, NY.

Cianciara A. Cianciara B., 2005. *Method of evaluation of mining tremors prediction on the basis of the analysis of asymmetry of seismoacoustic signals emission*. Arch. Min. Sci., Vol. 50, No 3, p. 317-326.

Cianciara A. Cianciara B., 2006. *The meaning of seismoacoustic emission for estimation of time of mining tremors occurrence*. Arch. Min. Sci., Vol. 51, No 4, p. 563-575.

Dubiński J., Konopko W., 2000. *Tąpania. Ocena, prognoza, zwalczanie*. GIG Katowice.

Gibowicz S. J., Kijko A., 1994. *An Introduction to Mining Seismology*. Academic Press, San Diego.

Gibowicz S.J., Lasocki S., 2001. *Seismicity induced by mining: Ten years later*. Adv. Geophys., 44, 39-181.

Głowacka E., 1993. *Application of the extracted volume as a measure of deformation for the seismic hazard evaluation in mines*. Tectonophysics, 202, 285-290.

Gołda I., Kornowski J., 2011. *Zastosowanie rozkładu Gutenberga-Richtera do prognozy zagrożenia sejsmicznego, wraz z oceną jego niepewności*. Górnictwo i Geologia, 6(3), p. 49-62

Hand D., Mannila H., Smyth P., 2001. *Eksploracja danych*. MIT Press, WNT.

Idziak A., Sagan G., Zuberek W.M., 1991. *An analysis of frequency distributions of shocks from the Upper Silesian Coal Basis*. Publ. Inst. Geophys. Pol. Acad. Sci., M-15(235), 163-182.

Idziak, A., Zuberek, W. M., 1995. *Fractal analysis of mining induced seismicity in the Upper Silesia Coal Basin*. [In:] Mechanics of Jointed and Faulted Rocks (H. P. Rossmanith, ed.). Balkema, Rotterdam, 1995, 679-682.

Jakubowski J., 2014. *A predictive model of daily seismic activity induced by mining developed with data mining methods*. Geoinformatica Polonica, 13,

Kabiesz J., 2008. *Badanie kategoryzacji zagrożenia tąpaniami z wykorzystaniem sieci neuronowych. Górnicze zagrożenia naturalne 2008*. Prace naukowe GIG. Górnictwo i Środowisko 7, GIG, Katowice.

Kornowski J., Kurzeja J., 2008. *Krótkookresowa prognoza zagrożenia sejsmicznego w górnictwie*. Główny Instytut Górnictwa, Katowice

Larose D.T., 2008. *Metody i modele eksploracji danych*. PWN, Warszawa.

Lasek M., 2002. *Data mining. Zastosowania w analizach i ocenach klientów bankowych*. BMiB, Warszawa.

Lasocki S., 1992. *Weibull distribution for time intervals between mining tremors*. Publ. Inst. Geophys. Polish Acad. Sci., M-16 (245), 241-260.

Lasocki S., 1993. *Statistical prediction of strong mining tremors*. Acta Geophys., Pol. 41, 1993, 1197-234.

Lasocki S., Orlecka-Sikora B., 2008. *Seismic hazard assessment under complex source size distribution of mining-induced seismicity*. Tectonophysics, 456, 2008, 28-37.

Leśniak A., Isakow Z., 2009. *Space-time clustering of seismic events and hazard assessment in the Zabrze-Bielszowice coal mine, Poland*. International Journal of Rock Mechanics & Mining Sciences, 46, 918-928.

Marcak H., 1993. *The use of pattern recognition method for predicting of the rockbursts*. [In:] R.R Young, (ed.) Rockbursts and Seismicity in Mines. Balkema, Rotterdam, 222-226.

Migut G., Harańczyk G., 2011. *Zastosowanie statystyki i data mining w badaniach naukowych i inne materiały*. StatSoft Polska, Kraków.

Orlecka-Sikora B., 2008. *Resampling methods for evaluating the uncertainty of the nonparametric magnitude distribution estimation in the Probabilistic Seismic Hazard Analysis*. Tectonophysics, 456, 38-51.

Sikora M., 2011. *Induction and pruning of classification rules for prediction of microseismic hazards in coal mines*. Expert Systems with Applications, 38, 6748-6758.

StatSoft, Inc., 2011. STATISTICA (data analysis software system), version 10.

Stec K., 2008. *Statystyczna zależność aktywności sejsmicznej górotworu od parametrów eksploatacji w kopalniach Górnośląskiego Zagłębia Węglowego*. Przegląd Górniczy, 64, 4, 26-34.

Tadeusiewicz R., Gąciarz T., Borowik B., Leper B., 2007. *Odkrywanie właściwości sieci neuronowych przy użyciu programów w języku C#.* Wydawnictwo PAU, Kraków.

Received: 07 May 2014