The Seismic Shock Prediction Algorithm Based on the Rock Mass Behaviour Data

TOMASZ MOROŃ1, BOŻENA STARUCH2, BOGDAN STARUCH2, SŁAWOMIR TOMASZEWSKI3, and AGNIESZKA WYŁOMANSKA4

1 Silesian University of Technology, Katowice, Poland
2 University of Warmia and Mazury in Olsztyn, Poland
3 University of Warsaw, Poland
4 Wrocław University of Science and Technology, Poland

(Communicated to MIIR on 31 October 2021)

Study Group: ESGI 144, 17–22 March, 2019 Polish Academy of Sciences, Warsaw

Communicated by: Kamil Kulesza

Industrial Partner: KGHM

Team Members: Kamil Kulesza5, Polish Academy of Sciences, Polish Academy of Sciences; Tomasz Moron, Silesian University of Technology; Bożena Staruch, University of Warmia and Mazury in Olsztyn; Bogdan Staruch, University of Warmia and Mazury in Olsztyn; Sławomir Tomaszewski, University of Warsaw Agnieszka Wyłomanska, Wrocław University of Science and Technology

Industrial Sector: Mining

Key Words: Seismicity, data analysis, mining

MSC2020 Codes: 62

Summary

KGHM S.A. exploits copper ore deposits in underground mining facilities. As a result of this operation the seismic activity of the rock mass is induced. One of the symptoms of seismic activity of the rock mass is the occurrence of high-energy seismic shocks. These phenomena can lead to severe destructions in mine workings. Resulting from that is a threat to work safety in the area of seismic shock occurrence and risk of damage to mine’s property. Particularly strong seismic shocks may also pose a threat to objects on the surface. The level of seismic activity of the rock mass depends on many factors that can be divided into:

• factors related to the environment in which the operation is carried out;
• factors related to methods of conducting the operation.

In the report authors propose an algorithm for prediction of the occurrence of seismic shocks with a given energy.

5 Corresponding contributor: kamil.kulesza@maths.com.pl
The seismic shock prediction algorithm based on the rock mass behaviour data

Contents

1 Introduction and problem formulation 4

2 Input data description 5

3 The proposed solution 8

4 Alternative approaches 15

5 Conclusions 18
The seismic shock prediction algorithm based on the rock mass behaviour data

Abstract

KGHM S.A. exploits copper ore deposits in underground mining facilities. As a result of this operation, the seismic activity of the rock mass is induced. One of the symptoms of seismic activity of the rock mass is the occurrence of high-energy seismic shocks. These phenomena can lead to severe destructions in mine workings. Resulting from that is a threat to work safely in the area of seismic shock occurrence and risk of damage to mine’s property. Particularly strong seismic shocks may also pose a threat to objects on the surface. The main objective of this project is to create an algorithm for the prediction of the occurrence of seismic shocks with a given energy. The secondary goal is as follows: based on the proposed algorithm, create a method for selection of operating parameters (e.g. blasting) leading to provoking (approaching and unavoidable seismic phenomena) in a given time.

1 Introduction and problem formulation

KGHM extracts and processes natural resources. The company possesses a geographically diversified portfolio of mining projects. It owns production plants on three continents - in Europe, South and in North America. Polish copper deposits - one of the biggest in the world - are exploited by three underground mines: "Lubin", "Polkowice-Sieroszowice" and "Rudna". On the other side of the ocean, KGHM owns six mines: Robinson, Carlota (USA), McCreedy West, Morrison (Canada) and Franke and Sierra Gorda (Chile). Apart from copper, these mines also produce molybdenum, nickel, gold, palladium, and platinum. KGHM ranks among the best producers of silver and copper in the world.

KGHM exploits copper ore deposits in underground mining facilities. As a result of this operation, the seismic activity of the rock mass is induced. In the seismic vibrations, one can observe seismic events which appear after the mining activity, like blasting, provoked relaxation of rock, and some unexpected events, like natural rock burst. One of the symptoms of seismic activity of the rock mass is the occurrence of high-energy seismic shocks. These phenomena can lead to severe destructions in mine workings. Resulting from that is a threat to work safely in the area of seismic shock occurrence and risk of damage to mine’s property. Particularly strong seismic shocks may also pose a threat to objects on the surface. Thus it is important to create algorithms for prediction of the occurrence of seismic shocks with a given energy. On the other side, it is also crucial to create a method for selection of operating parameters (e.g. blasting) leading to provoking (approaching and unavoidable seismic phenomena) in a given time.

Microseismic monitoring gives mining engineers information about the local state and stress conditions of the rock mass. By receiving real-time information on the location of seismic events, engineers and operators can identify where these events are occurring relative to
The seismic shock prediction algorithm based on the rock mass behaviour data

Mine openings and active workplaces and visualize how these conditions are changing over time. By understanding how the rock mass is behaving, operators can infer which mining activities are affecting the overall structure, and by how much. A challenge in microseismic monitoring is to correctly identify seismic events from blasts, as blasts contaminating the database have an adverse effect on any analysis, [3].

In the literature, one can consider different approaches of classification and identification of the seismic vibrations in order to recognize and analyze the sources of their different properties [2, 3, 11, 15]. As it was mentioned, the most important parameters that may influence the seismic activity are mining activity like blasting, provoked relaxation of rock or natural movements of the rock mass. However, there are additional parameters that should be taken under consideration, like conditions in the area of exploitation (geology, the direction of exploitation, the proximity of exploited areas) or information related to the exploitation (geometry of workings, the progress of exploitation front, blasting, etc.).

The main goal of this project is to propose the prediction system in order to predict the possible seismic activity in the considered prediction period. The system will be based on the data related to the seismic activity. More details of the input data we present in the next sections. Additional data that need to be taken under consideration are information of mining activity related to the seismic vibrations as well as information about the conditions in the area of exploitation or parameters of the exploitation. The input data should be defined by the experts from the mine who have the knowledge about sources of seismic vibrations and their relations with different parameters. In this report, we only present our vision of the considered issue and give the possible solution to the given problem.

The problem is formulated as follows: based on the given data from different sources, namely from the seismic activity (seismic vibration time series and their parameters), the information of the mining activity (like blasting) and the information of the area of explosion, create an algorithm for prediction of the occurrence of seismic shocks with a given energy. The additional goal is as follows: based on the proposed algorithm, create a method for selection of operating parameters (e.g. blasting) leading to provoking (approaching and unavoidable seismic phenomena) in a given time. The schematic formulation of the main problem is presented in Fig. 1.

2 Input data description

In the mine, through various monitoring systems, there are collected many data. The most important, from the current project point of view, are the data related to the rock mass activity (collected by the mine services like the seismic stations). However, in the mine, there are also collected data regarding conditions in the area of exploitation (geology, the
The seismic shock prediction algorithm based on the rock mass behaviour data

![Diagram](image)

Figure 1: The schematic description of the main problem.

direction of exploitation, the proximity of exploited areas, etc.) and data related to the exploitation (geometry of workings, the progress of exploitation front, blasting, etc.).

According to [9], where the real datasets from seismic vibrations in KGHM S.A. were also analyzed in the context of the seismic events detection and segmentation, the copper-ore underground mines in Poland use a prototype data acquisition system independent of the commercial seismic systems. In the system, the triaxial accelerometer is installed on the mining corridor roof (see Fig. 2). It measures the vibration acceleration in the range of $0.001100 \ m/s^2$ with a frequency range of $0.5400$ Hz. The sampling frequency is set to $1250$ Hz. A rigid connection between sensor and rock is ensured since the sensor is mounted using gypsum binder. Primarily, the sensor was located $20 \ m$ from the mining face and the distance was increasing up to $140 \ m$, due to the advance of mining works. The exemplary seismic vibration time series from the described system are presented in Figs. 3 and 4.

![Image](image)

Figure 2: The accelerometer located on the mining corridor roof, [9].

Except for the mentioned seismic vibrations in mines, there are collected and calculated (based on the seismic time series from different systems) additional parameters that describe and characterize the seismic vibrations. Those parameters are important in the context of the seismic activity prediction.

In [3] the authors indicated the importance of the following parameters in the seismic time
The seismic shock prediction algorithm based on the rock mass behaviour data

The exemplary real seismic signal with single events, [9].

The exemplary real seismic signal with multiple events, [9].

According to our knowledge, the most important parameters that characterize the seismic activity are: the energy of the seismic vibrations, time of duration of the seismic event, amplitude of the seismic event and P-wave onset moment. Thus, those parameters will be considered as the input data of the proposed prediction system. Some of the parameters are obtained directly from the monitoring systems in the mine, however, some of them need to be calculated. In the next sections, we indicate a few possible methods that can be applied in order to prepare the input parameters to the prediction system. The energy of the seismic vibrations seems to be the most important one, and its proper prediction may have huge practical importance. Thus, in the proposed prediction system, we concentrate on the prediction of the energy of the future seismic events. In order to the specific character of the energy distribution, in Fig. 5 we present the energy of seismic vibrations distributions for the exemplary region in the polish cooper-ore mine. As one can see, the probabil-
The seismic shock prediction algorithm based on the rock mass behaviour data

...ity of the occurrence of the small-energy events is much higher than the high-energy one. The high-energy events are the most dangerous and thus it is important to properly predict those seismic events. It is worth to highlight, the distribution of the energy differs for different areas of the mine. It is related to many factors, like parameters related to exploitation or geology.

Figure 5: The distribution of the energy of the exemplary area of the polish cooper-ore mine.

As it was mentioned in section 1, to the prediction system we will also give, as the input data, the additional parameters that are related to the seismic vibrations. Thus as the additional input data, we propose to give data regarding conditions in the area of exploitation (geology, the direction of exploitation, the proximity of exploited areas) and data related to the exploitation (geometry of workings, the progress of exploitation front, blasting, etc.). The input data should be clarified by the experts from the mine. At this stage, we only give the general idea of the input data for the proposed prediction system.

3 The proposed solution

In order to fulfill the main goal of the project we propose the prediction system. As it was mentioned in the previous section, as the first step, we prepare the data related to the seismic activity (seismic vibration time series and their parameters) as well as data that may have influence on the predicted values (exploitation data, data related to mining activity). In the Data pre-processing subsection we describe in details how the data should be prepared in order to apply them to the further analysis. Thus, proper data preparation is the second step of our algorithm. As a third step, we propose to use the similarity-based classifier in order to predict the energy of future seismic events. The details of this point we present in the subsections Parametrized family of the similarity-based classifier and Deeper analysis based on the classification results. Finally, we evaluate the prediction results. The details we present in the subsection Prediction evaluation. The block diagram of the proposed
The seismic shock prediction algorithm based on the rock mass behaviour data algorithm we present in Fig. 6.

**Figure 6:** The block diagram of the proposed algorithm.

### 3.1 Data pre-processing

As it was mentioned, as the first step we need to prepare the data to the further analysis. It is crucial especially in the context of the seismic vibration time series. As one can see in Figs. 3 and 4, the data have complicated structure, sometimes they are not complete or they are contaminated by different noises. Thus, in order to properly analyze the seismic time series and properly calculate their characteristic, there is a need to make the supplementing of missing data, removing outliers, and make the data cleansing (e.g. denoising). In signal processing, there are many methods that can be applied in the context of supplementing of missing data. Here we will not recommend any known methods. The algorithms depend on the real data and their specific behavior. In this report, we only indicate research papers where authors proposed the supplementing of missing data algorithms [5, 1, 18].

Very often the data need to be denoised because they may be contaminated by different sources. In recent years there appear many interesting papers devoted to this issue. We propose to apply the method presented in [17] because it is strictly devoted to the seismic time series analysis.

In order to calculate the characteristics of the seismic events, there is need to segment the time series (as in Figs. 3 and 4) in order to extract only the informative part from the time...
The seismic shock prediction algorithm based on the rock mass behaviour data series. In the literature, one can find different segmentation methods that can be useful in the context of seismic signals segmentation. We propose to apply the methods devoted strictly to the seismic time series. As we know, the seismic vibrations have a complicated structure, thus the segmentation algorithms should take under consideration this aspect. The proposed methods are described in details in [21, 2]

As the final step in the seismic time series preparation, we calculate the characteristics that can describe the seismic events. In further analysis, they are considered as the attributes \( \{a_1^1, \ldots, a_1^K, \ldots, a_L^1, \ldots, a_L^K\} \). The attributes should be defined by the experts from the mine. However, at this stage of the project, we propose to consider the following attributes related to the seismic time series: the energy of the seismic vibrations, time of duration of the seismic event, the amplitude of the seismic event and P-wave onset moment. They can be directly calculated from the extracted seismic events.

We assume the other input data related to the mining activity (attributes \( m_1, m_2, \ldots \)) are properly prepared and there is no need to make additional pre-processing. In the table below we present the exemplary information table that will be a base for the similarity-based classifier. The objects \( (ob_1, ob_2, \ldots) \) are the non-overlapping intervals of time with length \( i \) described in rows by the attributes values.

| Object | Sensor1 | Sensor2 | ...SensorL | Mine Activity |
|--------|---------|---------|------------|---------------|
| ob1    | ...     | ...     | ...        | ...           |
| ob2    | ...     | ...     | ...        | ...           |
| ob3    | ...     | ...     | ...        | ...           |
| ob4    | ...     | ...     | ...        | ...           |
| ob5    | ...     | ...     | ...        | ...           |

3.2 Parametrized family of similarity-based classifier

The solution is based on the family of classifiers \( Cl(t, E, T, r) \), where:

- \( t \) is an interval of time used to object description that depends on a set of parameters,
- \( E \) is a level of energy of the predicted event,
- \( T \) is a period for which we predict the energy level,
- \( r \) is a threshold that supports decision making.
The seismic shock prediction algorithm based on the rock mass behaviour data

For more information on classifiers based on coverings (similarity classes) see e.g. [19, 20]. It should be highlighted the parameters $t, E, T, r$ need to be discussed with the expert. They are strongly related to domain knowledge.

The proposed solution is described in details in the following parts. The schema of the algorithm is presented in Fig. 7.

![Decision Table Diagram](image)

**Figure 7:** The schema of the proposed solution.

### 3.3 Decision Table

The decision table (see Table 2) is an information table (see Table 1) extended with a decision attributes $D_{E,T}$, where $D_{E,T} = 1$ if an event with energy at least $E$ appeared during a period equal to $T$. The decision attribute can be understood as a subset of objects with a positive value of the decision.
The seismic shock prediction algorithm based on the rock mass behaviour data

Table 2: Decision Table

| Object  | Sensor1 | Sensor2 | ...SensorL | Mine Activity | $D_{E,T}$ |
|---------|---------|---------|------------|---------------|-----------|
| $ob_1$  | $a_1^1$ | $a_2^1$ | $a_3^1$    | $m_1$         | 0         |
| $ob_2$  | $a_1^2$ | $a_2^2$ | $a_3^2$    | $m_2$         | 0         |
| $ob_3$  | $a_1^3$ | $a_2^3$ | $a_3^3$    | $m_3$         | 1         |
| $ob_4$  | $a_1^4$ | $a_2^4$ | $a_3^4$    | $m_4$         | 1         |
| $ob_5$  |         |         |            |               |           |

3.4 Similarity Classes

The second step is grouping objects into similarity classes. The effect of such a grouping depends on attributes values and expert knowledge. Every similarity class is a subset of the set of objects, so it is represented as the binary column in the covering table (see Table 3). We assume that the similarity classes form a covering of the set of objects. If not, the entire set of objects is treated as an artificial similarity class.

Table 3: Covering Table

| Object | $C_1$ | $C_2$ | $C_3$ | ... |
|--------|-------|-------|-------|-----|
| $ob_1$ | 1     | 0     | 0     | ... |
| $ob_2$ | 0     | 1     | 0     | ... |
| $ob_3$ | 1     | 1     | 0     | ... |
| $ob_4$ | 1     | 1     | 1     | ... |
| $ob_5$ | ...   | ...   | ...   | ... |

3.5 Neighborhood of a new object and classification

Basing on attribute values of a new object $x$ we determine its neighborhood $N(x)$ as an intersection of all the similarity classes to which this object belongs.

The classification of the new object is executed by the parameter $r$ that should be determined by expert analysis. Then using a measure $\mu(N(x), D_{E,T})$ of inclusion of $N(x)$ to $D_{E,T}$ we are able to make the decision as follows:

$$D_{E,T}(x) = 1 \text{ if and only if } \mu(N(x), D_{E,T}) \geq r.$$  

In Fig. 8 we present the graphical representation of the relationship between neighborhood and the decision set.
Theseismicshockpredictionalgorithmbasedonthed rockmassbehaviourdata

Figure 8: The graphical representation of the relationship between neighborhood and the decision set.

In the literature [10, 16, 19] one can find different measures that can be useful here. Selection of appropriate measure depends on the application. Very often it is good to introduce a new measure which is the most appropriate to the considered problem. The most natural measure is as follows:

\[
\mu(N(x), D(E, T)) = \frac{\text{Card}(D_{E,T} \cap N(x))}{\text{Card}(N(x))}.
\]

3.6 More in-depth analysis of the classification results

In order to show how to analyze results of a classifier \( CL(i, E, T, r) \) let us assume that a probability (denoted as \( p \)) of prediction of decision "1" for the objects is already determined. Then we provide statistical analysis of a series of results obtained by the classifier for the sequences of objects. The analysis is based on the observation that in case the event has occurred in real life there is a series of \( \frac{T}{r} \) of ones in the column \( D_{E,T} \) starting from some object \( ob_0 \). The following example explains how the analysis of results of the classifier runs.

Example
Let \( ob_0 \) denote the first object with decision "1" and let \( n \leq \frac{T}{r} \) be a chosen series length (temporal window for observations). Then we can apply binomial distribution with parameters \( n \) and \( p \), where the success is to have the decision "1". Next, let us fix some confidence level \( c \) that is big enough to start some alert (more careful observations or another kind of actions). Now let \( n = 30, p = 0.7, c = 0.95 \). We ask the question what number of "1" (say \( k \)) in the temporal window indicates the high probability (at least 0.95) of occurring an event with
The seismic shock prediction algorithm based on the rock mass behaviour data energy at least \( E \). The answer is that the \( k \) is the 95th percentile. For data in this example \( k = 25 \).

If \( n \) is large enough, then the skew of the distribution is not too significant. In this case, a reasonable approximation to binomial distribution is given by the Gaussian distribution.

We can also provide a more in-depth analysis by executing a series of classifiers with different levels of energy. The cumulative information gained from such a series of result would be a tool for eliminating errors of the appropriate classifiers. As a final result, we obtain the predicted values of the energy for given prediction time \( T \).

3.7 Prediction evaluation

The model proposed during our work assumes that the feedback will be the predicted class of shocks. The high quality of predictions of energy classes should allow obtaining their distribution (let’s call it \( D_p \)-predicted distribution) with a similar shape to the distribution occurring from real data (call it \( D_r \)-real distribution). Firstly, to unify the obtained distributions, they should be normalized to the division \((0, 1)\). Then we can start the process of assessing the similarity of the obtained distributions.

The simplest approach proposed here is to determine the similarity coefficient \( S_{coff} \) as the square of Euclidean Distance \( E_D \) between normalized distributions:

\[
S_{coff} = E_D(N(D_p) - N(D_r))
\] (1)

![Figure 9: Visual presentation of proposed quality indicator.](image)

The total area under the obtained in Fig. 8 is the coefficient of the similarity between distributions (smaller value - greater similarity). The more advanced techniques that can be used to evaluate results can be methods used in areas such as processing time series. This method can be, for example, DTW (with various modifications), used successfully in
The seismic shock prediction algorithm based on the rock mass behaviour data [7] and processing biomedical data [8]. However, this approach requires tests on real data to choose the best assessment method. Based on the proposed metrics, a series of experiments evaluating the usability of the proposed solution will be possible.

4 Alternative approaches

4.1 Approach 1

As part of the current section, we propose a partial experiment based on the analysis of historical data. In the currently used system, the signals are divided into energy classes (on a scale of 0 – 8), based on the energy they release. Such a measurement is carried out for a time window containing an event whose location has previously been detected using practically used methods.

It can therefore be assumed that at the beginning of the experiment we have a signals, \( s_1, \ldots, s_J \), time points in these signals \( t_{j,1}, \ldots, t_{j,L} \), on the basis of which it is possible to determine windows \( w_{j,l} \) in which events are included and information about their energy class \( w_{j,l}, EC_{w_{j,l}} \). The experiment proceeds as shown in the Fig. 10. In the first phase, fragments will be extracted from the signals, on which, according to the methodology of KGHM (C), a classification based on the energy of events will be carried out. Each of the designated windows containing the signal fragment will be assigned an energy class label (D).

![Figure 10: Partial experiment scheme](image-url)

Figure 10: Partial experiment scheme
The seismic shock prediction algorithm based on the rock mass behaviour data

For each of the energy groups, an operation of grouping signals into subgroups will be carried out (it is suggest that 10 subgroups be created for each energy class) (E).

To determine the distance between the signal fragments, we suggest to use the Dynamic Time Warping method, which is used in the processing of biomedical signals and also gives good results as a component of the methods that group the time series [7, 8]. The distance measures will be used in the grouping method [12].

Thanks to this, each window will gain an additional label within its energy class $EC'$.

$$w_{j,l,EC(w_{j,l}),EC_j,l,EC(w_{j,l})}$$ (2)

Then a series consisting of energy classes in the form of pairs will be created for the whole signal (F):

$$[EC, EC']_{t_1}, [EC, EC']_{t_2}, [EC, EC']_{t_L}$$ (3)

Next step will be, based on the acquired time series, possible to choose a time series model and method of their prediction.

4.2 Approach 2

An alternative method of seismic shocks energy prediction can be developed based on the Generalized Linear Models (GLM) technique, introduced by Nelder and Wedderburn [13]. For the proofs of theoretical results see [6]. The GLM method is designed for building statistical models in situations, where the standard linear models approach cannot be applied, e.g. when we need to work with non-Gaussian distributions, or the regression function is non-linear. It has found successful applications in various fields, like modelling insurance claim counts or claim severities. In the case of seismic shock distribution the lack of Gaussianity is due to the non-negativity of energy, but also could be related to other issues.

The GLM method objective is to find a predictive distribution for a seismic shock energy, parameters of which depend on observed levels of influential factors, however similar construction can be employed to model directly the monetary costs related those events. It requires an assumption on the predictive distribution to belong to the exponential dispersion models (EDM) family, which is defined (in the continuous case) by having a probability density function of the following form:

$$f(y) = a(y, \phi) \exp \left\{ \frac{y\theta - b(\theta)}{\phi} \right\},$$

where $\theta$ and $\phi$ are parameters, $b(\theta)$ is a known function and $a(y, \phi)$ plays a role of a normalizing factor. In particular, the family contains not only the classical Gaussian distribution,
The seismic shock prediction algorithm based on the rock mass behaviour data but also the gamma and Tweedie distributions, which are proposed to be used for seismic shocks energy prediction model. Choosing the function $b(\theta) = -\log(-\theta)$ we get the gamma distribution.

Characterisation of the Tweedie distributions within the EDM class requires the concept of the variance function. It can be shown (see e.g. Ohlsson et al. [14]) that for a EDM random variable $Y$ we have $\mu = E(Y) = b'(\theta)$, (where $b'(\theta)$ denotes the derivative) and $\text{Var}(Y) = b''(\theta)\phi$. The variance function $v(\mu) = b''(b^{-1}(\mu))$ expresses dependence of variance on the mean. The Tweedie distribution is defined by having the power variance function:

$$v(\mu) = \mu^p$$

for some $p$. The Tweedie distribution with $p \geq 2$ is continuous and positive, which makes it appropriate for modelling shocks energy or cost values. Note, that the Tweedie family includes the Gaussian, gamma and inverse Gaussian distributions.

Suppose we have identified the relevant factors for seismic shocks energy prediction and defined an appropriate separation of each factor into some level classes. Then each observation falls into one cell described by some combination of factor levels. This can be expressed by the design matrix. For each cell $i = 1, \ldots, n$ we can form a vector $(x_{ij}), j = 1, \ldots, r$ of 0’s and 1’s indicating each factor level influencing the given cell. For a detailed description of a design matrix construction see [14].

The GLM method allows to estimate coefficients $\beta_j$ of a model of a form:

$$g(\mu_i) = \sum_{j=1}^{r} x_{ij} \beta_j,$$

where $\mu_i$ is the expected value of the predicted variable in cell $i$, and the link function $g$ is some monotone function. An important example of a link function is the logarithm, which is used for multiplicative models.

Parameter estimation can be performed using the maximum likelihood (ML) method. The ML equations have a form (see [14] for a derivation):

$$\sum_i w_i \frac{y_i - \mu_i}{v(\mu_i)g'(\mu_i)} x_{ij} = 0, \quad j = 1, \ldots, r$$

where $y_i$ is the observed value of the predicted variable in cell $i$, $w_i$ expresses the exposure (i.e. the number of observations in the cell in the case of energy prediction) and $\mu_i = \mu_i(\beta) = g^{-1}(\sum_{j=1}^{r} x_{ij} \beta_j)$ is a function of the vector of estimated coefficients. This system can be solved numerically, many standard statistical packages have built-in functions that perform necessary computations, see e.g. [4].

The GLM method proved its usefulness in many applications, like insurance pricing. We can see it is also well suited for the seismic shocks energy prediction problem.
5 Conclusions

In this report, we have considered the problem of prediction of future seismic events and their energy. The considered problem is extremely important from a practical point of view. Based on the predicted energy one can expect the dangerous events in the underground mine. The main goal of this project was to create an algorithm for the prediction of the occurrence of seismic shocks with a given energy. We have proposed the algorithm based on the similarity-based classifier. This method belongs to the data mining group of algorithms and seems to be very effective in the considered problem. As the input data, we propose to apply the parameters of the seismic vibration time series that characterize the seismic events. Additional parameters are related to the conditions in the area of exploitation (geology, the direction of exploitation, the proximity of exploited areas) and data related to the exploitation (geometry of workings, the progress of exploitation front, blasting, etc.). The input data should be defined by the experts and in the proposed algorithm we only present our idea of possible input data. As the output data, we obtain the predicted energy of the seismic events for a given predicted period. In our opinion, the proposed algorithm may be very effective. We proposed also the prediction evaluation algorithm that gives us information about prediction effectiveness. Finally, we demonstrate the alternative approaches that may be applied in the considered problem. We hope, the proposed idea may be used in the near future in the KGHM mines.

References

[1] V. D. Agostino, A. Zelenka, Supplementing solar radiation network data by co-Kriging with satellite images, International Journal of Climatology 12, 749-761, 1992.

[2] L.-J. Dong, X.-B. Li, G.-N. Xie, Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive bayes classification, Abstr. Appl. Anal. 2014, 459137, 2014.

[3] L.-J. Dong, J. Wesseloo, Y. Potvin, X.-B. Li, Discriminant models of blasts and seismic events in mine seismology, International Journal of Rock Mechanics & Mining Sciences 66, 282-291, 2013.

[4] P.K. Dunn, G.K. Smyth, Generalized Linear Models With Examples in R, Springer Science+Business Media, New York, 2018

[5] M. Gasior, Ł. Skowron, Methods of imputation of missing values and their influence on the results of segmentation research, Econometrics 4, 61-71, 2016.

[6] B. Joergensen, The Theory of Dispersion Models, Chapman & Hall, London (1997)
The seismic shock prediction algorithm based on the rock mass behaviour data

[7] M. Kotas, J. Leski, T. Moroń, J. G. Guzman, Hierarchical Agglomerative Clustering of Time-Warped Series. In International Conference on Man-Machine Interactions (pp. 207-216). Springer, Cham, 2017.

[8] M. Kotas, T. Moroń, ECG signals reconstruction in subbands for noise suppression, Biocybernetics and Biomedical Engineering 37.3, 453-465, 2017.

[9] D. Kucharczyk, A. Wyłomańska, J. Obuchowski, R. Zimroz, M. Madziarz, Stochastic modelling as a tool for seismic signals segmentation, Shock and Vibration, vol. 2016, Article ID 8453426, 13 pages, 2016, doi:10.1155/2016/8453426.

[10] A. Kumar, M. Banerjee, Definable and rough sets in covering-based approximation spaces. In: Li, T., Nguyen, H.S., Wang, G., Grzymala-Busse, J., Janicki, R., Hassanien, A.E., Yu, H. (eds.) RSKT 2012, LNCS 7414, Springer, Heidelberg, 488-495, 2012.

[11] H. Langer, S. Falsaperla, T. Powell, G. Thompson, Automatic classification and aposteriori analysis of seismic event identification at Soufriere Hills volcano, Montserrat, Journal of Volcanology and Geothermal Research 153, 1-10, 2006.

[12] J. M. Leski, M. Kotas, T. Moroń. PCA Based Hierarchical Clustering with Planar Segments as Prototypes and Maximum Density Linkage. Man-Machine Interactions 4. Springer, Cham, 2016. 507-516.

[13] J.A. Nelder, R.W.M. Wedderburn, Generalized linear models, J. R. Stat. Soc. Ser. A 135, 370-384 (1972)

[14] E. Ohlsson, B. Johansson, Non-Life Insurance Pricing with Generalized Linear Models, Springer-Verlag, Berlin

[15] C. P. Please, A. A. Wheeler, P. Wilmott, A mathematical model of cliff blasting, SIAM J. Appl. Math. 47, 117-127, 1987.

[16] P. Pagliani, Covering Rough Sets and Formal Topology. A Uniform Approach Through Intensional and Extensional Constructors. Transactions on rough sets, Springer, 109-145, 2016.

[17] M. Polak, J. Obuchowski, A. Wyłomańska, R. Zimroz, Seismic signal enhancement via AR filtering and spatial time-frequency denoising. In: Chaari F., Leskow J., Napolitano A., Zimroz R., Wyłomanska A. (eds) Cyclostationarity: Theory and Methods III. Applied Condition Monitoring, vol 6. Springer, Cham, 51-68, 2017.

[18] I. Pratama, A. E. Permansari, I. Ardiyanto, R. Indrayani, A review of missing values handling methods on time-series data, 2016 International Conference on Information Technology Systems and Innovation (ICITSI), 2016.

[19] B. Staruch, Classification model based on topological approximation space, LNCS, vol. 10314 LNAI, 570-578, 2017.
The seismic shock prediction algorithm based on the rock mass behaviour data

[20] B. Staruch, B. Staruch, A topological approximation space based on open sets of topology generated by coverings, LNCS, vol. 10314 LNAI, 130-137, 2017.

[21] R. Zimroz, M. Madziarz, G. Zak, A. Wyłomańska, J. Obuchowski, Seismic signal segmentation procedure using time-frequency decomposition and statistical modelling, Journal of Vibroengineering 17(6), 3111-3121, 2015.