GEOINFORMATICA POLONICA

13 : 2014

CONTACT – ADRES REDAKCJI

Wydział Geodezji Górniczej i Inżynierii Środowiska AGH
30-059 Kraków, al. Mickiewicza 30
tel. (012) 617-22-84
e-mail: maciasze@agh.edu.pl

POLSKA AKADEMIA UMIEJĘTNOŚCI
PRACE KOMISJI GEOINFORMATYKI

GEOINFORMATICA POLONICA

13

KRAKÓW 2014

Redaktor tomu
Jadwiga MACIASZEK

ISSN: 1642-2511

Złożono do druku: czerwiec 2014
Druk ukończono: październik 2014

Dystrybucja
Polska Akademia Umiejętności
ul. Sławkowska 17, 31-016 Kraków
e-mail: wydawnictwo@pau.krakow.pl

www.pau.krakow.pl

Skład komputerowy
FALL, ul. Garczyńskiego 2, 31-524 Kraków

Objętość: ark. wyd. 8; ark. druk. 12,25, nakład: 150 egz.

CONTENTS

Papers

SPIS TREŚCI

Artykuły

GEOINFORMATICA POLONICA
13: 2014

JACEK JAKUBOWSKI [1]
jacek.jakubowski@agh.edu.pl

A PREDICTIVE MODEL OF DAILY SEISMIC ACTIVITY INDUCED BY MINING, DEVELOPED WITH DATA MINING METHODS

Key words:

Seismicity induced by mining, induced seismicity, mining tremors, seismic hazard, rockburst hazard, longwall mining, data mining, predictive models, classification models, boosted trees, neural networks, logistic regression

Abstract

The article presents the development and evaluation of a predictive classification model of daily seismic energy emissions induced by longwall mining in sector XVI of the Piast coal mine in Poland. The model uses data on tremor energy, basic characteristics of the longwall face and mined output in this sector over the period from July 1987 to March 2011. The predicted binary variable is the occurrence of a daily sum of tremor seismic energies in a longwall that is greater than or equal to the threshold value of 10^5 J. Three data mining analytical methods were applied: logistic regression, neural networks, and stochastic gradient boosted trees. The boosted trees model was chosen as the best for the purposes of the prediction. The validation sample results showed its good predictive capability, taking the complex nature of the phenomenon into account. This may indicate the applied model's suitability for a sequential, short-term prediction of mining induced seismic activity.

MODEL PREDYKCYJNY DOBOWEJ AKTYWNOŚCI SEJSMICZNEJ INDUKOWANEJ EKSPLOATACJĄ, ZBUDOWANY METODAMI DATA MINING

Słowa kluczowe:

Sejsmiczność indukowana eksploatacją, wstrząsy górnicze, hazard sejsmiczny, zagrożenie tąpaniami, eksploatacja ścianowa, data mining, modele prognostyczne, modele klasyfikacyjne, drzewa wzmacniane, sieci neuronowe, regresja logistyczna

Abstrakt

W artykule przedstawiono budowę i ocenę predykcyjnego modelu klasyfikacyjnego dobowej energii sejsmicznej indukowanej eksploatacją ścianową węgla w partii XVI kopalni Piast. Model wykorzystuje dane o energii wstrząsów

[1] AGH University of Science and Technology, Department of Geomechanics, Civil Eng. and Geotechnics, Krakow, Poland

oraz podstawowe dane charakteryzujące front eksploatacji i wydobycie w tej partii w okresie od lipca 1987 do marca 2011. Dwustanową zmienną prognozowaną jest wystąpienie dobowej sumy energii wstrząsów w ścianie większej lub równej wartości progowej 10^5 J. Zastosowano trzy metody analityczne data mining: regresję logistyczną, sieci neuronowe i drzewa wzmacniane. Jako najlepszy dla celów prognozy wybrano model drzew wzmacnianych. Wyniki na próbie walidacyjnej pokazują jego wysoką zdolność predykcyjną, biorąc pod uwagę złożoność zjawiska. Wskazywać to może na przydatność zastosowanego modelu do konstrukcji bieżącej, krótkookresowej prognozy zagrożenia sejsmicznego.

1. Introduction

Rock mass intact at large depths is found in a state of primary, stable mechanical equilibrium, shaped by the weight of the overlying rock mass, by geological history as well as local tectonic structures. Mining activity disturbs the primary mechanical equilibrium of rock mass, evokes rock mass movements, stress and elastic energy concentrations. Local breaks and sudden discharges of elastic strain energy take place, along with the propagation of seismic waves, i.e. tremors. If this coincides with excavation destruction, the dynamic phenomenon is called a rockburst.

Rockburst hazard is among the most serious natural hazards related to mining activity. The method of mining seismology is one of the primary methods for assessing the rockburst hazard. It is based on continuous monitoring of tremors induced by mining activity and on observation of the intensity of this seismic activity and its variations.

Due to its relationship with rockbursts, seismic activity induced by mining is a subject of much interest to geophysicists and miners. The most important and most difficult research problems include prediction of the size, time of occurrence, and location of tremors as well as the level of seismic activity in a given region based on seismological historical data.

Prediction most often pertain to seismic hazard, i.e. the probability that a tremor of a specific size will occur at a specific time and in a specific region. If the seismic hazard is estimated in moving time windows, then it is a function of time (t). To estimate it, the Gutenberg-Richter relation and the variations of its parameters (Gibowicz 1979, Lasocki 1993, Trifu et al. 1997, Gołda and Kornowski 2011, Kijko et al. 1987), other probability distributions (Lasocki 1992; Kijko 1997, Cianciara et al. 2005), including extreme value distributions (Epstein and Lomnitz 1966, Idziak et al. 1991), as well as non-parametric estimators of probability distributions (Kijko et al. 2001, Lasocki and Orlecka-Sikora 2008, Orlecka-Sikora 2008) are applied. Many other methods are also used, e.g. a fractal description of energy, time, and the locations of tremor centers as well as the variation of the fractal dimension (Xie 1993, Idziak and Zuberek 1995), space-time cluster analysis (Leśniak and Isakow 2009), dependency of seismic energy and mined out rock volume (McGarr 1976, Głowacka 1993, Vallejos and McKinnon 2011), pattern recognition (Marcak 1993), neural networks (Kabiesz 2008, Rudajev and Ciz 1999, Jakubowski and Tajduś 2014), boosted trees (Jakubowski and Tajduś 2014), rules induction methods (Sikora 2011), time series analysis considering the sum of seismic and acoustic emission energy (Kornowski and Kurzeja 2010).

Comprehensive reviews of applied methods and achievements in this scope can be found, e.g. in publications by Gibowicz and Lasocki (2001), Gibowicz (2009), Gibowicz and Kijko (1994), Dubiński and Konopko (2000).

Most seismic hazard prediction methods are methods using the classical approach of mathematical statistics based on probability distributions and empirical relations. Based on observations of seismic activity over the time interval $(t - \Delta T, t)$, a seismic hazard is predicted in the interval $(t, t + \Delta t)$. The quantities predicted are: seismic energy, magnitude, seismic moment, or time between the consecutive tremor events.

The approach to predicting induced seismic activity shown in this article is the classification data mining approach applying analytical machine learning methods (boosted trees, neural networks) and statistical methods (logistic regression). The prediction is the result of developing of a multivariate model describing the relation

of input (predictors) and output (response) variable. Only variables with values in the interval (t, $t + \Delta t$) that are known at instant t can be predictors. In particular, the parameters of mining longwalls, seismic energy in previous periods, mining output in previous periods, and mining output scheduled for the prediction period (t, $t + \Delta t$) are such variables.

The proposed model combines historical data on seismic activity from the catalog of tremors with basic information on mining output and longwall faces (Jakubowski et al. 2014). The response variable is the occur- rence of a daily sum of seismic energy in a longwall $\geq 10^5$ J. Thus, a seismic hazard, which is a continuous variable, is not being predicted here, but one of the seismic intensity measures converted to a binary variable.

2. Data mining vs. classical statistics approach

According to Cabena et al. (1998), data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large databases. It refers to the use of various methods that can automatically extract information from data with little user intervention. All definitions include some element of technology of data analysis and databases (Berry and Linoff 2000, Seidman 2000, Hand and Mannila 2001, Larose 2008, Lasek 2002, Migut and Harańczyk 2011, Ratner 2003, Koronacki and Ćwik 2005).

Data mining can be divided into predictive (supervised) data mining and descriptive (unsupervised) data mining. Predictive data mining solves problems of regression, classification, and time series analysis. Descriptive or structure discovery data mining includes cluster analysis, sequence, association link analysis and dimension reduction problems.

The analytical methods used in data mining are e.g. classification and regression trees, neural networks, stochastic gradient boosted trees, random forest, support vector machines (SVM), multivariate adaptive regression splines, naive Bayes, logistic regression, discriminant analysis, k-means, k-nearest neighbors, principal component analysis (PCA), independent component analysis (ICA), descriptive statistics and graphs.

The features that distinguish the classical statistical approach to data analysis and the data mining approach from each other are of analytical and technological character (Tab. 1).

The first criterion is the manner of evaluating models and estimating uncertainty through classical statistical methods and data mining. Classical statistical methods use mathematics to evaluate models and estimate uncertainty in a strict parametric manner. As a result, generalization to the entire population is strict. This comes at the cost of rigorous assumptions and requirements which the data must fulfill for such a generalization to be true.

Data mining uses data that does not usually fulfill these theoretical assumptions and requirements. Quantitative model evaluation is not strict, but empirical. Because of this, generalizations to the entire population are not strict but still may be evaluated quantitatively.

Empirical model and uncertainty evaluation in the data mining scheme is based on a division of the data set into training, testing, and validation sets. The model is built on the training set with the involvement of the testing set and is evaluated using the validation set (Fig. 1).

In the classical statistical approach, the form of the relation between studied variables is usually assumed in advance, for example, based on a known mechanism relating the studied variables or empirical knowledge. The model development is based only on an estimation of its parameters.

Data mining models are usually non-parametric, black box models, built only on the basis of data. This is why these models are useful in predicting but of small use in explaining, and they usually lack a direct interpretation.

Technological criteria distinguishing classical statistics approach and data mining, are related, among others, to data sources, data transformations, processing, implementation, and the user environment (Tab. 1).

Table 1. Analytical and technological criteria distinguishing classical statistics approach and data mining
Tab. 1. Analityczne i techniczne kryteria odróżniające klasyczne podejście statystyczne i data mining

	Criterion	Classical statistics	Data mining
Analytical criteria	Model and uncertainty evaluation	Mathematical	Empirical
	Models	Parametric models	Black box models
	Sample	Random sample, small	Encountered sample, large
	Main goal	Explanation, confirmation	Prediction, new knowledge
Technological criteria	Data sources	Data in a single, flat file, usually small	Data from multiple sources. Integration or data warehouse necessary
	Data transformations	Data is ordered or does not require automatic transformations	Automatic aggregation, cleaning, and complex transformations are necessary
	Processing	Relatively exhaustive operations per record/variable	Sequential processing. Scalability is key
	Implementation	Does not require implementation or a large IT infrastructure	Implementation in databases as PMML, XML, SQL, C, VB code
	User environment	Conventional user environment	Data mining work-space

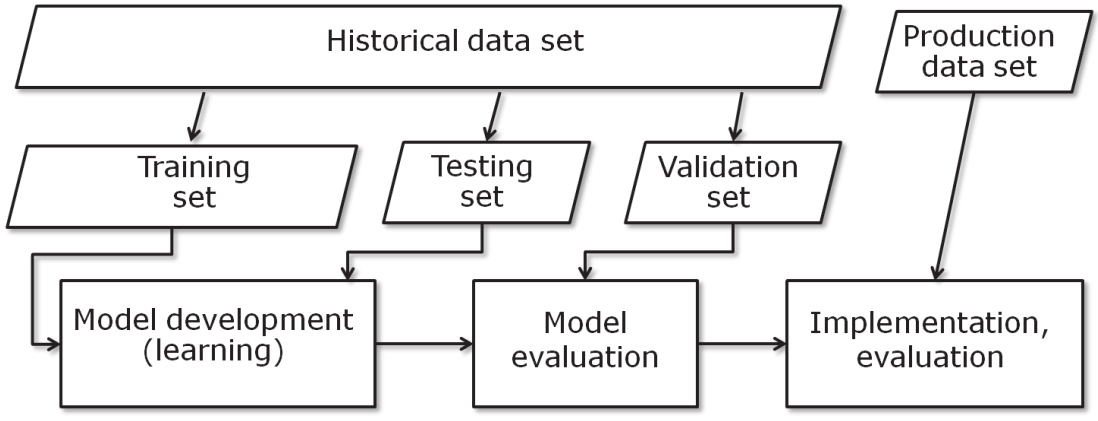


Fig. 1. Data mining model development and evaluation, including uncertainty evaluation
Rys. 1. Budowa modelu oraz ocena modelu i niepewności w metodyce data mining

3. Seismic activity induced by mining in sector XVI of the Piast mine

As a result of mining activity, rock mass's failure and sudden discharges of elastic energy are observed as tremors. Sometimes, they lead to a rockburst: a sudden destruction of excavations, lining, or machines, that may also result in fatalities.

The rockburst hazard is present in deep mines in the RSA and Canada, in the USA, China, India, Russia, and the Czech Republic. In Poland, the rockburst hazard is present both in coal mines and in the copper basin.

Tremors of the rock mass create a seismic wave which is received by seismometers and geophones in the form of vibrations. The geophysical interpretation of these signals allows for the location of hypocenters, estimation of seismic energy, seismic moment and other source parameters. On this basis, seismological criteria of rockburst hazards are established, and methods of rockburst prevention are developed.

The Piast mine lies in the eastern part of the Upper Silesian Coal Basin and has a mining area of 48 km^2. It has an output of about 4.5 mln tons of coal per year, mined exclusively in the longwall system with roof caving. Coal seams of series 200 (Łaziska layers, Westphalian C) are mined.

The mine's geophysics station conducts continuous observation of tremors. Such observations constitute a source of input data for the predictive models shown further in the paper.

In sector XVI of the Piast mine, four coal seams located at a depth of 500–800 m were exploited: 205, 206, 207 and 209. Due to the lack of historical data concerning mined output in deposit 206, analysis was performed for deposits 205, 207, and 209.

4. Input data, problem definition, and the scheme of model development

Input data includes data from the seismic catalog as well as basic data concerning daily output from longwalls running in coal seams 205, 207, and 209 in sector XVI of KWK Piast, periodically over the years 1987–2011.

Data from the seismic catalog includes 2400 tremors attributed to longwalls running in sector XVI, the date and time of the event, seismic energy and hypocenter coordinates. Data on mining is, above all, daily output and basic longwall parameters: length, height, seam number. The data accounts for pauses in output caused by days off work. Long technological breaks and periods during which mining in sector XVI was not carried out, were not taken into account. In total, the data included 2759 days. Among those days, there are days in which there was no output and days in which there were no tremors.

Data on tremors, excavated output and longwalls was collected independently in separate files, thus it was extracted, aggregated and aligned by days. If on a given day, two adjacent longwalls were running, then it was accepted that aggregated longwall output is the sum of output from both longwalls, the aggregated length of the longwall is the sum of the lengths of all longwalls, and the aggregated height of the longwall is the mean height of both longwalls running that day.

The prediction problem was defined as follows: On a given day, will the sum of tremor seismic energy released in a longwall be greater than or equal to 10^5 J ? This is a predictive and classification problem, and data mining techniques were used to solve it. For this problem, a binary response variable equal to 1 was defined if the sum of seismic energy emitted in a longwall during the day $\geq 10^5$ and equal to 0 if the energy was lower. The predictors in the model are variables that describe: daily output, daily output on the previous day, longwall length, longwall height, the integer part of the logarithm of the sum of tremor seismic energies on the previous day, two days earlier, and three days earlier. Thus, the model consists of seven continuous predictors and a binary response variable.

The analytical scheme of the construction and evaluation of component models and implementation of the final model has been presented on Figures 2 and 3. All the models have been built with STATISTICA Data Miner v.10 (StatSoft, Inc. 2011). The first stage of analysis is a preliminary data overview, data cleansing, handling missing data, definition of a response (dependent) variable according to the stated predictive problem. The selection of predictors (independent variables) is a key

operation which is usually performed in the process of multiple analysis and model evaluation.

The data set is divided into a learning set, on which models are developed, and a validation set, on which they are evaluated. Models were built using three completely different analytical methods, all using the same learning data set. Developed models were then run on the validation set, evaluated, and compared.

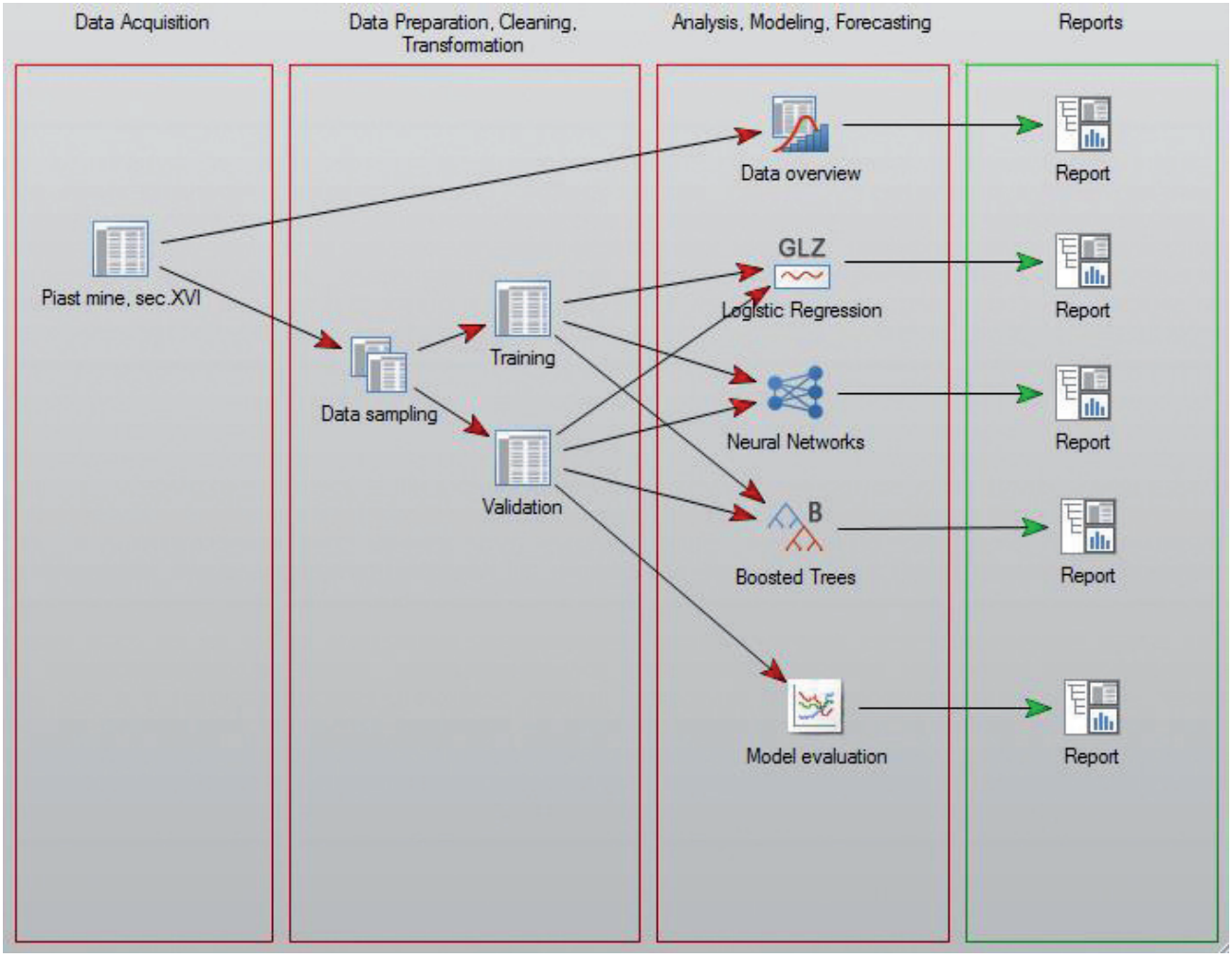


Fig. 2. Analytical scheme of the development and evaluation of models in the data mining work-space
Rys. 2. Schemat analityczny budowy i oceny modeli w przestrzeni roboczej data mining

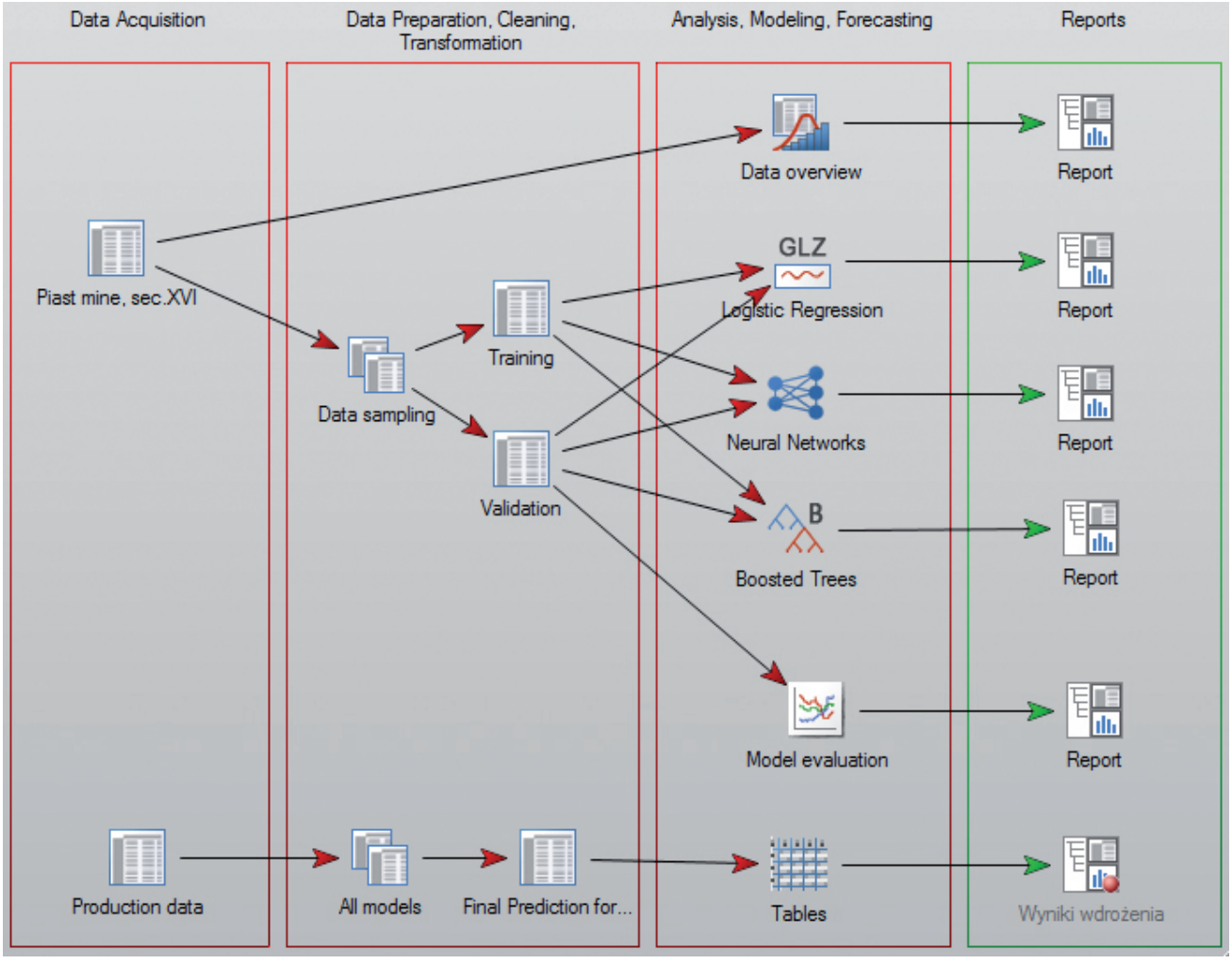


Fig. 3. Analytical scheme of the implementation of models in the data mining work-space in which they were developed

Rys. 3. Schemat wdrożenia analityczny budowy i oceny modeli w przestrzeni roboczej data mining, w której zbudowano model

5. Applied analytical methods and model parameters

Three analytical methods were used to develop the predictive models: logistic regression, neural networks (multi-layer perceptron), and stochastic gradient boosted trees.

All component models were built using the STATISTICA Data Miner v.10 specialized system.

In logistic regression, the probability model is a logistic function (Fig. 4). Its shape, a smoothed step, is commonly used in classification models. For the developed predictive model, this is the probability of the occurrence of a daily sum of seismic energy in a longwall which is greater than or equal to the threshold value 10^5 J.

$$P(Y=1)=\frac{e^{\beta_0+\beta_1 x_1+....+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+....+\beta_k x_k}} \quad (1)$$

where $P(Y=1)$ is the probability that the response variable will take the value of 1, x_1 to x_k are independent variables, and β_0 to β_k are coefficients of the logistic regression equation.

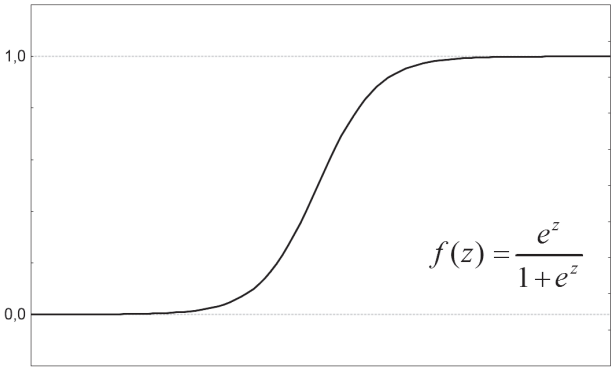


Fig. 4. Shape of the logistic function
Rys. 4. Kształt funkcji logistycznej

Coefficients in the logistic regression equation are estimated by the maximum likelihood method, the model's significance can be evaluated with the LR test (likelihood ratio), and the significance of regression equation coefficients with the Wald test. It is apparent that logistic regression is a statistical method, however, it will be applied further not in the classical statistical approach, but according to data mining approach.

The second applied analytical method was neural networks, specifically a multilayer perceptron (Tadeusiewicz et al. 2007). This network is trained under supervision and consists of input, hidden, and output layer. Neurons have a linear aggregating function. Every neuron on a layer is connected to every neuron of the neighboring layers, and a weight is assigned to every connection. Training has the purpose of finding a weight vector, for which the network's output values are closest to the observed ones.

The process of neural network training consists of many successive stages (epochs), during which a one-time pass through the entire training set is performed, on the basis of which the weight vector is modified and error is minimized. The learning error is minimized on the training set and monitored on the test sample in order to avoid over-learning of the network. Initial weight values are chosen at random. During learning, the local minimum of the learning error is determined, which is why, even for networks with the same structure and trained with the same data, the result of learning may be different.

Among the several hundred semi-automatically trained and compared networks, a network with the following parameters was selected: 6 neurons in one hidden layer (Fig. 5), learning using the BFGS method with 17 epochs, entropy error function, hyperbolic tangent neuron activation function for the hidden layer and Softmax for the output layer (StatSoft, Inc. 2011).

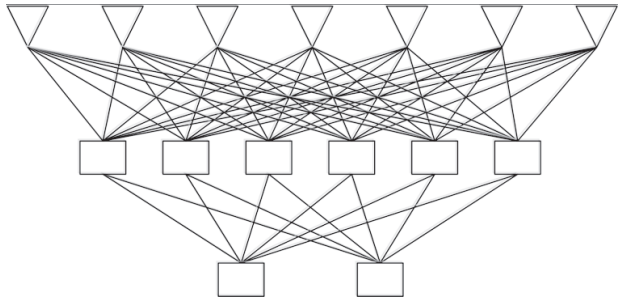


Fig. 5. The multilayer perceptron scheme corresponding to the applied network: 7 input neurons, 6 neurons in the hidden layer, and two output neurons
Rys. 5. Schemat perceptronu wielowarstwowego odpowiadającego zastosowanej sieci: 7 neuronów wejściowych, 6 neuronów w warstwie ukrytej i dwa neurony wyjściowe

BFGS, the applied learning method, is similar to the quasi-Newton method of finding the minimum of a function, in this case, the minimum of the error function in the space of weights. A certain approximation of the inverse Hessian matrix, determined by means of the Broyden, Fletcher, Goldfarb, Shanno formula, is used for this purpose (StatSoft, Inc. 2011).

The third applied analytical method was the stochastic gradient boosted trees method. The boosted trees method consists of many simple CART type trees. Component trees entering into the model are generated itera-

tively. The first tree is a tree built with training data. The next tree is created on training data which is modified with weights in such a way that erroneously classified cases received 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 average response of all component trees. Model training, or generation of component trees, takes place on the training set. The process of adding new component trees to the model is stopped when the model error (mean polynomial deviation) on the test sample ceases to decrease.

The model developed for the purposes of this analysis and used later consists of 132 component trees, each with 5 nodes. Figure 6 shows as example, two of the component trees.

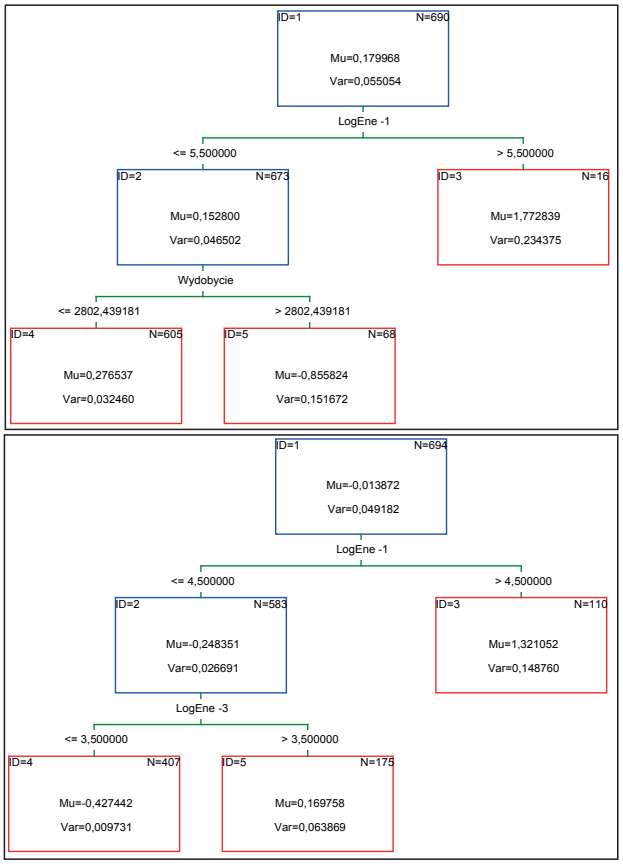


Fig. 6. Example of two component trees in the boosted trees model

Rys. 6. Dwa przykładowe drzewa składowe modelu drzew wzmacnianych

6. Evaluation of models using the validation sample

For empirical evaluation of data mining classification models, there are used, among others ROC curves and lift charts.

Lift charts illustrate to what degree the prediction is better than a random selection of cases. For example, for the neural network model in Figure 7, if 10% of all days are to be predicted as high seismic energy days (10% at horizontal axis) and the lift value is 5.0 (vertical axis, curve for neural networks), it means that the fraction of days with the actually observed high seismic energy in the prediction is five times greater than on average in the entire data set.

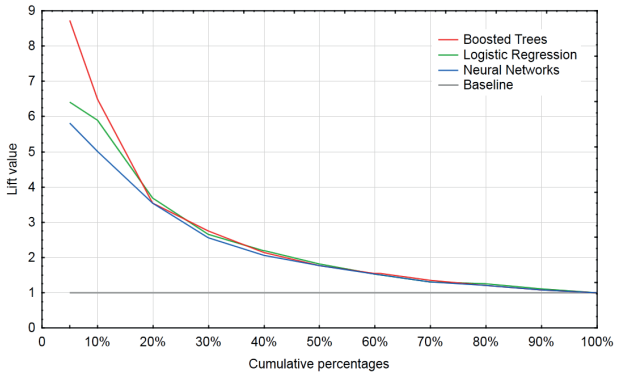


Fig. 7. Lift chart for boosted trees, logistic regression, and neural network models

Rys. 7. Wykres przyrostu dla modeli drzew wzmacnianych, regresji logistycznej i sieci neuronowych

A ROC chart allows to evaluate the predictive power of models and to compare models performance on a validation set (Harańczyk and Stępień 2008). Predictive power is expressed by the area under the ROC curve (AUC). AUC equal to 1 indicates a perfect model, and AUC equal to 0.5 indicates a random model. ROC charts for the three developed models are shown on Figure 8. The boosted trees model has the greatest AUC, equal to 0.847. A value of 0.847 indicates that the model reflects the studied phenomenon well. AUC is one of the more popular criteria for general evaluation of models.

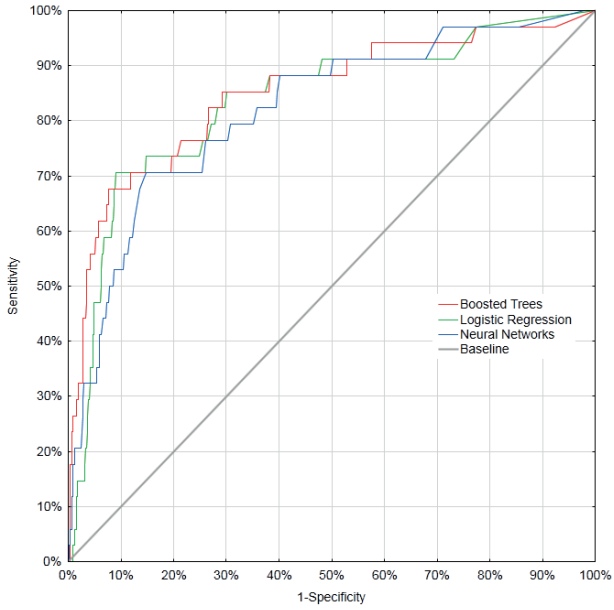


Fig. 8. ROC chart for boosted trees, logistic regression, and neural network models

Rys. 8. Wykres ROC dla modeli drzew wzmacnianych, regresji logistycznej i sieci neuronowej

It was decided, that the fraction of days with high energy was equal to 7%, which is equivalent to about 25 days with high energy per year. Cut-off points for each model result from the above prediction operation condition. It was also decided that the final model will be the best of the three models, meeting the condition above. According to this criterion, the boosted trees model is the best model (Tab. 2, Fig. 7). The boosted trees model is also the best model according to the criterion of the area under the ROC curve (AUC) as it was noted above.

Table 3 shows the classification matrix for the final boosted trees model. Observations are shown in columns and predictions are shown in rows. The validation sample size was equal 812. Out of which, there were 778 days with a sum of seismic energy in a longwall of less than 10^5 J and 34 days with a sum greater than or equal to this value. The prediction indicated 57 days with high energy, out of which 19 days (33.3%) were predicted correctly (true positive) and 38 days incorrectly (false positive). The prediction indicated 755 days with low energy, out of which 15 days were predicted incorrectly (false negative) and 740 days correctly (true negative).

Table 2. Model evaluation criteria: local (ACC, PPV, sensitivity) and global (AUC)

Tab. 2. Oceny modeli: lokalne dla wybranego warunku prognozy (ACC, PPC, czułość) oraz ocena ogólna (AUC)

Model	Accuracy (ACC)	Positive predictive value (PPV)	Sensitivity	Area under ROC curve (AUC)
Boosted trees	93,5%	33,3%	55,9%	0,847
Logistic regression	92,7%	28,1%	47,1%	0,838
Neural networks	91,7%	21,1%	35,3%	0,819

Table 3. Classification matrix for the final boosted trees model. Columns contain observations and rows contain predictions

Tab. 3. Macierz klasyfikacji dla modelu końcowego drzew wzmacnianych. W kolumnach obserwacje, w wierszach prognoza

	Observations 1	Observations 0	Observations Total	Fraction of 1 in total observations	Fraction 1	Fraction 0	Fraction Total
Prediction 1	19	38	57	33,3%	55,9%	4,9%	7,0%
Prediction 0	15	740	755	2,0%	44,1%	95,1%	93,0%
Prediction Total	34	778	812	4,2%	100,0%	100,0%	100,0%

Among the 34 days with actually observed high energy, 19 were indicated by the prediction (55.9%). In total, the prediction was correct for 93.5% of days and incorrect for 6.5% of days. In other words, the accuracy of the decision rule (ACC) amounted to 93.5%, and the accuracy of high energy predictions (PPV) amounted to 33.3%, the sensitivity of the prediction was 55.9%, and the specificity of the prediction was 95.1%. The lift value for the prediction was equal to 8.0. See Bradley (1997) for the definitions of te above mentioned specificity, sensitivity, PPV, ACC, AUC and other classification model characteristics. Figure 9 shows classification histogram.

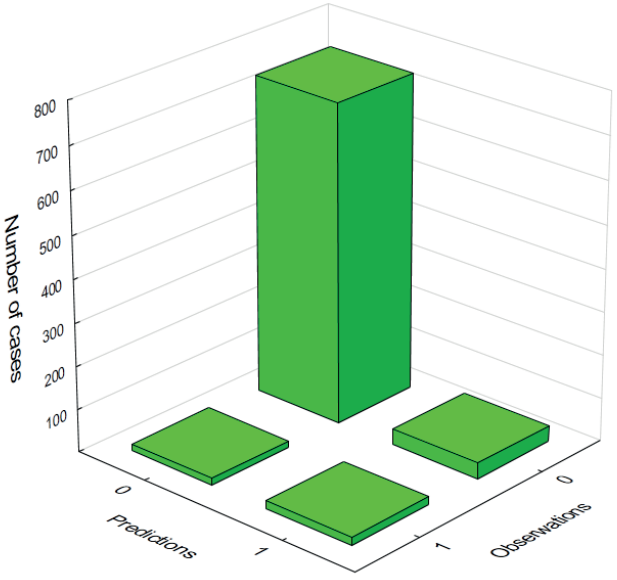


Fig. 9. Classification histogram reflecting classification matrix

Rys. 9. Histogram klasyfikacji dla macierzy klasyfikacji

The developed and evaluated predictive model can now be implemented, that is, applied for practical purposes, for new data that was unavailable at the time of model development, referred as production data. A tremor and longwall output database which is updated on a daily basis makes it possible to perform a sequential, daily predictions for the next day in the case of a predictive problem defined as it was here.

7. Conclusion

Most methods of predicting seismic activity induced by mining based on seismological data are seismic hazard estimation methods that use the classical approach of mathematical statistics, based on probability distributions and empirical relations.

The approach to predicting mining induced seismic activity shown in this article is the classification data mining approach applying analytical machine learning methods (boosted trees, neural networks) and statistical methods (logistic regression). Here, one of the measures of the intensity of induced seismic activity is predicted, not the seismic hazard itself. The shape of the probability distribution, local stationarity, or tremor independence are not assumed. Probability distributions of energy and other source size parameters are not studied as in classical, statistical estimation of the seismic hazard. The prediction is the result of development of a non-parametric multivariate model based only on data that describes non-linear relations of predictors and the response variable, but without assuming the form of these relations.

The data mining approach allows many predictor variables, does not put strict assumptions for data and error distributions and tolerates interactions and dependency of variables. It does, however, require a large sample. Instead of strict mathematical evaluation of the statistical model and its uncertainty, the quality of which is strongly dependent on the assumptions and simplifications of the model, the data mining approach provides an empirical evaluation of model that is performed on a validation sample after the model has been built.

Three models were built: logistic regression, neural network, and boosted trees, and the final model was the best of these three models. In terms of the adopted criterion, the stochastic gradient boosted trees model proved to be the best one.

The prediction included 55.9% of all days with a high energy observed in the validation sample. In the group of days with high energy indicated in the prediction, 33.3% were indicated correctly. The accuracy of the decision rule reached 93.5%. Considering the particularly complex and stochastic nature of the phenomenon, these numbers showed a good predictive capability of

the model and the utility of the presented approach for a sequential, short-term prediction of mining induced seismic activity.

Acknowledgements

The author is grateful to Łukasz Lenart and Łukasz Ożóg for collecting raw data used for this work. The author wish to thank the management and employees of Piast colliery for their support in collecting data necessary for the research.

Prepared for the AGH UST research project 11.11.100.197

References

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

Bradley A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition, 30 (7), pp. 1145–1159.

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., Takuska-Węgrzyn E. 2005. Methods of tremor risk monitoring based on analysis of non-homogeneity of stream of registered seismoacoustic emission. Geoinformatica Polonica, 7:2005.

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

Epstein B., Lomnitz C. 1966. A model for the occurrence of large earthquakes. Nature 211, pp. 954–956.

Gibowicz S.J. 1979. Space and time variations of the frequency-magnitude relation for mining tremors in the Szombierki coal mine in Upper Silesia, Poland. Acta Geophys. Pol. 27, pp. 39–50.

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, pp. 39–181.

Głowacka E. 1992. Application of the extracted volume as a measure of deformation for the seismic hazard evaluation in mines. Tectonophysics 202, pp. 285–290.

Gołda I., Kornowski J. 2011. Zagrożenie sejsmiczne i sejsmiczne zagrożenie tąpnięciem. Część II. Ocena ilościowa i analiza jej niepewności. Prace naukowe GIG. Górnicze zagrożenia naturalne 2008. Górnictwo i Środowisko 4/2/2011 GIG, Katowice.

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

Harańczyk G., Stępień M. 2008. Ilustrowana sztuka podejmowania decyzji. Matematyka Społeczeństwo Nauczanie, 41, pp. 12–15.

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) pp. 163–182.

Jakubowski J., Lenart Ł., Ożóg Ł. 2014: Predykcyjny model dobowej emisji energii sejsmicznej indukowanej eksploatacją górniczą. Przegląd Górniczy 70(3), 18–25.

Jakubowski J., Tajduś A. 2014. Predictive regression models of monthly seismic energy emissions induced by longwall mining. Archives of Mining Sciences 59(3).

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

Kijko A. 1997. Seismic hazard assessment in mines. [In:] S.J. Gibowicz and S. Lasocki (eds.) Rockbursts and Seismicity in Mines. Balkema, Rotterdam, pp. 247–256.

Kijko A., Drzezla B., Stankiewicz T. 1987. Bimodal character of the distribution of extreme seismic

events in Polish mines. Acta Geophys. Pol. 35, pp. 157–166.

Kijko A., Lasocki S., Graham G., Retief S.J.P. 2001. Non-parametric seismic hazard analysis in mines. [In:] van Aswegen, G., Durrheim, R.J., Ortlepp, W.D. (eds.), Rockbursts and Seismicity in Mines Dynamic Rock Mass Response to Mining. South African Inst. Min. Metal, Johannesburg, pp. 493–500.

Kornowski J., Kurzeja J. 2010. Energetyczna prognoza zagrożenia sejsmicznego. Prace naukowe GIG, Górnicze zagrożenia naturalne. Górnictwo i Środowisko 4/3/2010 GIG, Katowice.

Koronacki J., Ćwik J. 2005. Statystyczne systemy uczące się. Wydawnictwa Naukowo-Techniczne, Warszawa.

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), pp. 241–260.

Lasocki S. 1993. Statistical prediction of strong mining tremors. Acta Geophys. Pol. 41, pp. 197–234.

Lasocki S., Orlecka-Sikora B. 2008. Seismic hazard assessment under complex source size distribution of mining-induced seismicity. Tectonophysics 456, pp. 8–37.

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, pp. 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, pp. 222–226.

McGarr A. 1976. Seismic moments and volume changes. J. Geophys. Res. 81, pp. 1487–1494.

Migut G., Harańczyk G., 2011. Zastosowanie statystyki i data mining w badaniach naukowych i inne materiały, http://www.statsoft.pl/czytelnia/.

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

Ratner B. 2003. Statistical Modellingand Analysis for Database Marketing – Effective Techniques for Mining Big Data. Chapman & Hall / CRC.

Rudajev V., Ciz R. 1999. Estimation of mining tremor occurrence by using neural networks. Pure Appl. Geophys. 154, pp. 57–72.

Seidman C. 2000. Zgłębianie i analiza danych w Microsoft SQL Server 2000. Microsoft.

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

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

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

Trifu C.I., Shumila V., Urbancic T.I. 1997. Space-time analysis of microseismicity and its potential for estimating seismic hazard in mines. [In:] S.J. Gibowicz and S. Lasocki, (eds.) Rockbursts and Seismicity in Mines. Balkema, Rotterdam, pp. 295–298.

Vallejos J.A., McKinnon S.D. 2011. Correlations between mining and seismicity for re-entry protocol development. International Journal of Rock Mechanics & MiningSciences 48, pp. 616–625.

Xie H. 1993. Fractals in Rock Mechanics. Geomechanics Research series, Balkema.

GEOINFORMATICA POLONICA
13: 2014

GRZEGORZ LENDA[1]
grzenda@agh.edu.pl

USING SPLINE FUNCTIONS FOR THE SHAPE DESCRIPTION OF THE SURFACE OF SHELL STRUCTURES

Key words:

shell structures, quadrics, splines, approximation

Abstract

The assessment of the cover shape of shell structures makes an important issue both from the point of view of safety, as well as functionality of the construction. The most numerous group among this type of constructions are objects having the shape of a quadric (cooling towers, tanks with gas and liquids, radio-telescope dishes etc.). The material from observation of these objects (point sets), collected during periodic measurements is usually converted into a continuous form in the process of approximation, with the use of the quadric surface. The created models, are then applied in the assessment of the deformation of surface in the given period of time. Such a procedure has, however, some significant limitations. The approximation with the use of quadrics, allows the determination of basic dimensions and location of the construction, however it results in ideal objects, not providing any information on local surface deformations. They can only be defined by comparison of the model with the point set of observations. If the periodic measurements are carried out in independent, separate points, then it will be impossible to define the existing deformations directly. The second problem results from the one-equation character of the ideal approximation model. Real deformations of the object change its basic parameters, inter alia the lengths of half-axis of main quadrics. The third problem appears when the construction is not a quadric; no information on the equation describing its shape is available either. Accepting wrong kind of approximation function, causes the creation of a model of large deviations from the observed points.

All the mentioned above inconveniences can be avoided by applying splines to the shape description of the surface of shell structures. The use of the function of this type, however, comes across other types of limitations. This study deals with the above subject, presenting several methods allowing the increase of accuracy and decrease of the time of the modelling with the splines.

WYKORZYSTANIE FUNKCJI SKLEJANYCH DO OPISU KSZTAŁTU POWIERZCHNI BUDOWLI POWŁOKOWYCH

Słowa kluczowe:

budowle powłokowe, funkcje sklejane, aproksymacja

[1] AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering. This work has been made within scientific research program no. 11.11.150.005

Abstrakt

Ocena kształtu płaszcza budowli powłokowych stanowi istotne zagadnienie zarówno z punktu widzenia bezpieczeństwa, jak i funkcjonalności konstrukcji. Najliczniejszą grupę wśród tego typu budowli stanowią obiekty posiadające kształt kwadryk (chłodnie kominowe, zbiorniki gazów i płynów, czasze radioteleskopów itp.). Zebrany podczas cyklicznych pomiarów kontrolnych punktowy materiał obserwacyjny, zamieniany jest najczęściej na postać ciągłą w procesie aproksymacji, przy użyciu powierzchni stopnia drugiego. Wyznaczone modele, służą następnie do oceny zaistniałych w czasie deformacji powierzchni. Postępowanie takie niesie ze sobą jednak kilka istotnych ograniczeń. Aproksymacja przy użyciu kwadryk, pozwala na wyznaczenie podstawowych wymiarów i usytuowania budowli, jednak jej efektem są obiekty idealne, nie niosące żadnych informacji o lokalnych deformacjach powierzchni. Można je określić dopiero przez porównanie modelu z punktowym zbiorem obserwacji. Jeżeli pomiary okresowe nie były prowadzone w tych samych punktach, wówczas nie ma możliwości bezpośredniego określenia zaistniałych deformacji. Drugi problem wynika z charakteru jednorównaniowego, aproksymacyjnego modelu idealnego. Rzeczywiste deformacje obiektu zmieniają podstawowe jego parametry, m.in. długości półosi głównych kwadryki. Trzeci problem pojawia się, gdy budowla nie jest kwadryką, nie ma również informacji o równaniu opisującym jej kształt. Przyjęcie mylnego rodzaju funkcji aproksymującej, spowoduje wyznaczenie modelu o dużych odchyłkach od obserwowanych punktów.

Wszystkie wymienione niedogodności można ominąć, stosując do opisu kształtu powierzchni budowli powłokowych funkcje sklejane. Użycie tego typu funkcji napotyka jednak na innego typu ograniczenia. Niniejsze opracowanie omawia powyższą tematykę, przedstawiając kilka metod pozwalających na zwiększenie dokładności i szybkości modelowania za pomocą funkcji sklejanych.

1. Introduction

Shell structures make a group of objects more and more common in many areas of the construction industry. The best known and commonest are the objects having the shape of the quadrics (second degree surface). They can usually be found in industry as objects of the shape of spheres, cylinders or rotating hyperboloids (gas and liquid tanks, cooling towers). Quadrics are also present in civil engineering, where they may have a more advanced form, resulting not only from the function of the construction, but also its aesthetic values. The design of objects of higher shell complexity require the application of multi-equation surfaces. In the simplest case, these can be surfaces made by spanning the given curve alongside the other curve, i.e. the conducting curve. Spanning can also take place in case of a higher number of curves and be connected with change of the scale, which leads to very elaborated models. The most sophisticated shapes of the surface can be obtained based on splines, especially NURBS functions. Examples of the construction of shell structures were presented in figure 1.

The need for periodic control of the surface shape of shell structures can be caused by different factors. The most important ones are connected with stability of construction, which is particularly important in case of self-supporting constructions. Deformations of shell surface can significantly weaken the durability of construction. Some objects (e.g. radio-telescopes) require preservation of required parameters of shape due to their special functions. The first stage in the control process of all the types of shell objects is a measurement the set of point data of proper density and accuracy. Nowadays it is carried out mainly with the application of the reflectorless polar method, with greater and greater participation of laser scanning [Shan and Toth 2009, Tsakiri et al. 2006]. The subsequent stage involves the replacement of a point dataset with a continuous model of surface, allowing further analyses. In the case of objects of the shape of quadrics, a standard procedure is LSM approximation with the help of the surface equation of second degree (Ahn 2004, Gocał 2010). This allows the determination of general equation of surface, and then of the basic features of

Fig. 1. Examples of shell structures. (source: www.trendhunter.com/trends/dalian-shell-museum, www.gc.shuttle.de/gc/limbach-o/fvu/gruppe03/radioteleskop.jpeg, Geodimex – company materials, own materials, http://www.oplista.pl/sites/default/files/business/zlote_tarasy.jpg)

Rys. 1. Przykłady obiektów powłokowych. (źródło: www.trendhunter.com/trends/dalian-shell-museum, www.gc.shuttle.de/gc/limbach-o/fvu/gruppe03/radioteleskop.jpeg, Geodimex – company materials, own materials, http://www.oplista.pl/sites/default/files/business/zlote_tarasy.jpg)

the surface: coordinates of the centre, lengths of main half-axes and their orientation towards the axis of the system of coordinates. Modelling of more complex surfaces requires the application of multi-equation mathematical representation of the object. This task can be fulfilled by splines (Farin 2002, Kiciak 2000), allowing the obtaining of a smooth description of the object preserving the continuity of the curvature. These functions can be also useful in the description of the surface of classical quadrics, especially when the construction has numerous and chaotically distributed deformations of the shell (Fig. 2).

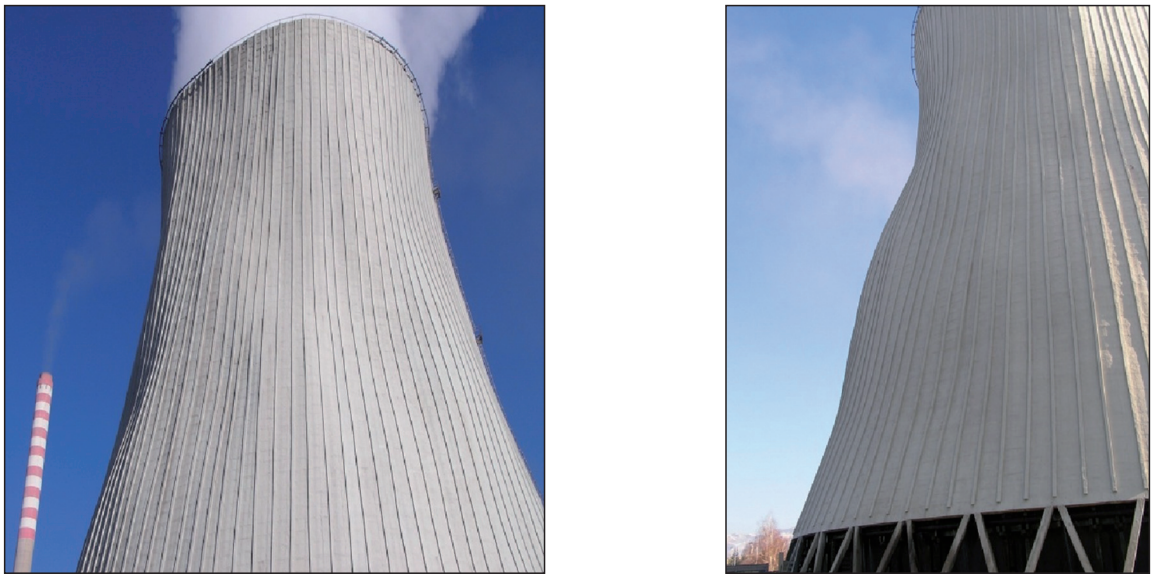
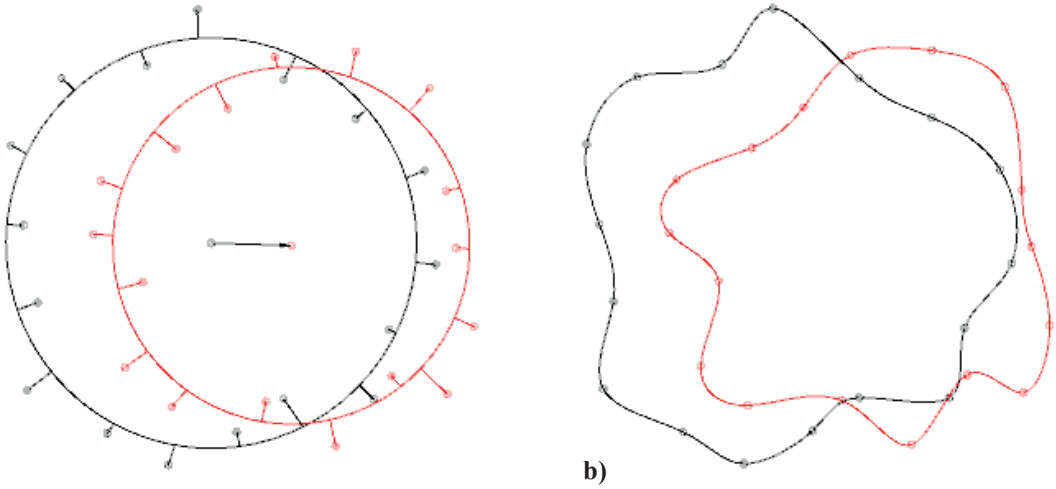


Fig. 2. Deformations of shell of cooling tower (Bitola, Macedonia). (source: Geodimex – company materials)
Rys. 2. Deformacje płaszcza chłodni kominowej (Bitola, Macedonia). (źródło: Geodimex – company materials)

Generally, approximation models of surface can be divided into two groups: one-equation (described with one equation of surface, e.g. a quadric) and multi-equation (splines). Each of the approximation methods has its characteristic features, which were presented in figure 3 and table 1.



a) **b)**

Fig. 3. Double (periodical) description of the object, which changed its shape in time: one-equation approximation, b) splines. (source: own materials)
Rys. 3. Podwójny (okresowy) opis obiektu, który zmienił w czasie swój kształt: aproksymacja jednorównaniowa, b) funkcje sklejane

Table 1. Comparison of the features of modelling with one-equation approximation and splines
Tab. 1. Porównanie cech modelowania za pomocą aproksymacji jednorównaniowej i funkcji sklejanych

One-equation approximation	Splines
Advantages of splines	
Ideal model of surface, with no information on local deformations	Describe local deformations of surface
In case of two periodic measurements and observations made in different points, it is impossible to directly compare deformations of the objects	One can compare deformations in any points, not necessarily the measured ones
The change of the situation in one point changes the whole model	The change of the situation in one point changes a small part of the model
It is necessary to know the equation of surface, according to which the model was constructed. In the case of the complex objects and lack of project documentation it can be impossible	One can describe any shapes with a standard, uniform spline model
The disadvantages of splines	
Small sensitivity to the dispersion of the measurement data	Errors of the model appearing with the description of objects measured with small density and regularity of the distribution points on the surface
Quick model construction	Slow construction and processing the surface for large sets of points

Despite the fact that splines make a widely applied tool for the description of the shape of shells, they have limitations, which may in certain cases significantly complicate the modelling. In the further part of the paper, selected, characteristic limitations of spline functions are presented as well as several useful modelling techniques, allowing, to certain extent, the elimination of the mentioned disadvantages.

2. Factors limiting the accuracy of approximations obtained with spline functions

Quadrics are a mathematic equivalent of steel rod bent within the range of elasticity. The curvature of the rod is connected with the bending moment in the Bernoullie-Euler equation (Ahlberg et al. 1967, Kosma 1999):

$$K = \frac{f''(x)}{[1+(f'(x)^{2}]^{3/2}} = \frac{M_z}{JE} \tag{1}$$

where:
K – curvature,
M_z – bending moment,
E –Young module,
J – inertia moment.

Relation (1) is used to derive the applied in practice splines, however it requires the acceptance of certain simplification. When analysing the rod shape in the area of elasticity, one can assume its small bending (tangent to the horizontally established rod will be almost horizontal). Thus one can assumed the value of the first derivative present in the denominator of the equation describing the curvature, as equalling zero. This assumption facilitates the calculations very much, because spline function can then be determined as a result of solving the system of

linear equations. This is particularly important also in terms of the speed of the calculations, because the spline function between every pair of points is described with a separate equation. Solving system *n–1* of a non-linear equations (for *n* points), requires more calculations than it takes place in case of the system of linear equations.

In practice, splines are used to describe the shape of objects of far more complicated than a bent elastic rod, which involves unavoidable errors in approximation. These errors, to much extent, depend on the provided measurement material. Irregular distribution of the points on the object surface can make a serious obstacle, preventing from achieving a good accuracy of approximation. In the case of objects measured with too small density, an additional problem can be a usually desired property of quadrics, meaning the formation of a model of minimal total curvature (Ahlberg et al. 1967, De Boor 1978). Measurement data of small density make the model flatter. Figures 4.a–c present several models illustrating the described issues.

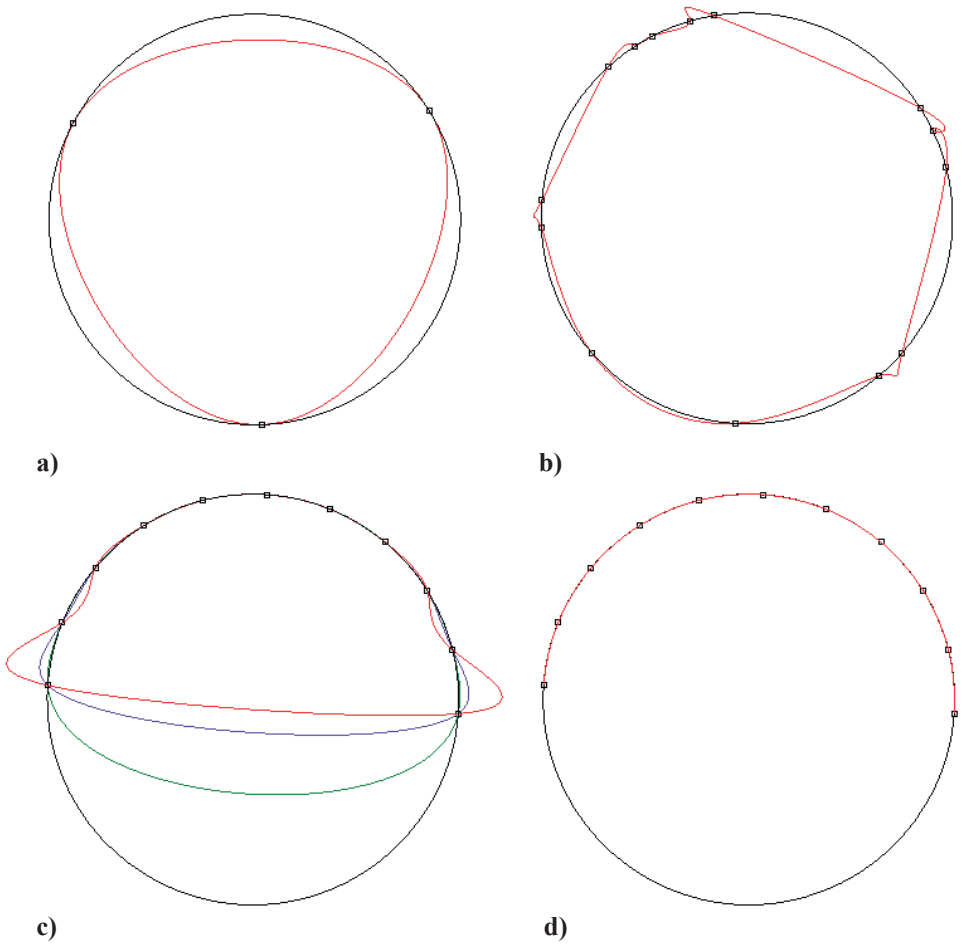


Fig. 4. Splines describing models: a) measured with too small density, b) measured with too small regularity, c) immeasurable in the bottom part, treated as a closed object, d) immeasurable in the bottom part, treated as an open object. (source: own materials)

Rys. 4. Funkcje sklejane opisujące model: a) mierzony ze zbyt małą gęstością, b) mierzony ze zbyt małą regularnością, c) niemierzalny w dolnej części, traktowany jako obiekt zamknięty, d) niemierzalny w dolnej części, traktowany jako obiekt otwarty

Models presented in figures 4.a and b show separately problems referring to density and regularity of distribution points, model 4.c makes in a certain meaning their combination. In the case of treating this model as closed, there are no measurement points in the bottom part. Due to the minimization of the spline curvature, it makes the function flatter. This involves waves on the neighbouring sections. Irregularity of the observations distribution results from a significant distance between extreme bottom points and points on remaining sections, leads to additional enhancement of waving. Figure 4.c shows several varieties of splines, with the application of different kinds of parameterization, which will be discussed in the further part of the text. Model 4.d, identical in terms of the observation material with model 4.c, was treated as an open object. Thus it is described only on upper section, covered with points, which decisively improves accuracy of approximations. The data set is in this case both dense enough and regular. In the chapter 3 presented a few selected, standard techniques for creating and modifying spline functions, which in a broader sense are presented in the work (Lenda 2006).

3. The methods of making and modify splines

The most common form of splines are B-spline functions (Farin 2002, Kiciak 2000). They are determined by linear combination of base polynomials N_i, of coefficients in the form of control points d_i.

$$S_i(t) = \sum_{i=0}^{n-m-1} d_i N_i^m(t) \qquad t_i = \{t_0,..,t_n\} \tag{2}$$

where:
t_i – knots,
d_i – control points,
N_i – base polynomials defined by recurrence Mansfield-de Boor-Cox formula (Farin 2002),
m – the degree of function,
n – the number of points

Making spline functions can be accompanied by different modification methods, which let to achieve the desired accuracy of approximations. The most important were presented below.

3.1. Parameterization

The first basic factor allowing the influence on the shape of the function is the application of a proper parameterization (Farin 2002, Haron et al. 2012, Lee 1989, Lim 1999), establishing knots (arguments) t_i of function, in the following way:

$$\begin{aligned} &t_0 = 0 \\ &t_j = t_{j-1} + T_j \qquad \text{for} \quad t \in [0, t_{\max}] \\ &t_n = t_{n-1} + T_n = t_{\max} \end{aligned} \tag{3}$$

where:
T_i – parameters

Parameters T_i are established in practice the most common in three ways:

a) parameterization with a uniform distance:

$$T_i = const \tag{4}$$

b) parameterization with the distance between points:

$$T_i = l_{i,i-1} = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{5}$$

c) parameterization with the square root of distance between points:

$$T_i = \sqrt{l_{i,i-1}} \tag{6}$$

where:
$l_{i,\,i-1}$ – distance between points

The fist of these does not regard in any case mutual distribution of points, leading to the worst effects of approximations (Fig. 4.c red). This is the least complicated method, which positively influences the time of calculations in the case of processing very large point sets. The parameterization with the distance between points is recommended to the description of objects of mild changes in shape (Fig. 4.c green). The parameterization with the square root of distance (Fig. 4.c purple) brings the best effects at the description of objects changing their shape quickly. There are also other, more complex types of parameterization, e.g. Foley's (Farin 2002) or Universal (Haron et al. 2012), allowing in some cases achieve better accuracy approximations; however they are not commonly implemented in the software used in the modelling with splines.

The application of a proper parameterization has a paramount meaning in achieving a desired accuracy of approximations, however it cannot (like other methods), fully compensate failures resulting from the lack of a proper distribution of the observation on the object.

3.2. NURBS Functions

The often encountered expansion of B-spline functions are NURBS functions (Non Uniform Rational B-splines) (Farin 2002, Hearn and Baker 1997):

$$\mathbf{S}(t) = \frac{\sum_{i=0}^{n-m-1} w_i \mathbf{d}_i N_i^m(t)}{\sum_{i=0}^{n-m-1} w_i N_i^m(t)} \tag{7}$$

where:
w_i – weights

The introduction of weights for given points, allows to local modelling of the curve. This process, carried out manually, is very inefficient, because it is difficult to determine the values of weights, causing respective changes. These changes, consequently, cause deformations of the functions on neighbouring sections, which requires the introduction of counter-weights for them. Thus, practical significance have only the automatic methods which are looking for the optimal solution (e.g. of the smallest total curvature). Such search is carried out by iterations (Dan and Lancheng 2006, Safraz 2008) and for large point sets, require significant calculation efforts, which in practice makes them not very useful.

3.3. Smoothing approximation

This kind of approximation allows to plot a spline function not by the given points, but by their equivalents, regarding the value of the measurement error and determined by a given level of smoothing the function. The determination of parameters of the approximation spline is carried out in the process of the minimization of the following functional (Diercx 1995, Farin 2002):

$$r\sum_{i=1}^{n}\left(\frac{\mathbf{p_i} - \mathbf{S}(t_i)}{\boldsymbol{\sigma}_{\mathbf{p_i}}}\right)^2 + (1-r)\int_{t_1}^{t_n}\left(\mathbf{S}''(t)\right)^2 dt \tag{8}$$

where:
$\mathbf{p}_i$ – coordinates x, y or z point,
$\mathbf{S}(t_i)$ – spline function,
$\boldsymbol{\sigma}_{p_i}$ – coordinate error,
r – approximation coefficient,
t – independent variable of the function in the parametric form

Expression (8) consists of two segments, the first of which is responsible for the LSM approximation, while the second for the total curvature of the function. Operating the smoothing parameter r within the borders [0,1], allow the function to larger pressure on good smoothing or good adjustment to the observation set. One should also bear in mind that the application of the approximation, i.e. „releasing" the function of the given points always flattens the model. This is connected with the already mentioned property of splines, which assume minimum energy state (minimum total curvature of function).

3.4. Approximation limiting the number of segments of spline function

The previous discussion focused on the first of the limitations of spline functions, listed in Table 1, connected with the errors of approximation, resulting from the distribution of the measurement points. Second limitation is connected with a relatively large amount of calculations, necessary in the construction of models, based on sets of very numerous points. The problem can be solved applying the approximation, changing the sets of points on the given sections, on individual segments of a spline function (Diercx 1995, Farin 2002).

$$\sum_{t=1}^{r}\left[\mathbf{p_j} - \sum_{i=0}^{k}\mathbf{d_i} N_i^3(t)\right]^2 \rightarrow \min \tag{9}$$

where:
r – the number of given points $\mathrm{p_j}$ used in approximation,
k – the number of control points of the spline, $k = m - 2$,

where:

m – the number of knots, which for approximation function can be at most: $m < r + 4$

This way one can describe complex structures of points, including the ones originating from the laser scanning. With very dense sets of points, regarding the amount of calculations, however it is necessary to filter the input data. The choice of point sets included in single segment of spline is usually automatic and corresponds the arbitrary division resulting from the uniform parameterization. Application of spline functions to process the point clouds originating from the laser scanning, is used only in the latest years and it is now a continuous development.

4. Making spline surfaces

The presented so far methods of making and modifying spline functions referred to curves, however they are concern surfaces on the similar principles. Spline surfaces are formed with the method of lofting. In its first step, based on the given points, curves $\mathbf{S}_i(t)$ are formed, in the second, on the curves the surface is spanned. The procedure of forming the surface is similar to determining curves, with that difference, that the „points" correspond to previously created curves, and the „curve" corresponds to surface $\mathbf{S}_{i,j}(t,u)$. Spline surfaces are described as tensor cross-product $\mathbf{S} = \mathbf{F} \otimes \mathbf{V}$ of the space of base functions: $N_i^3(t) \in \mathbf{F}$ and $N_j^3(u) \in \mathbf{V}$ (Farin 2002, Kiciak 2000):

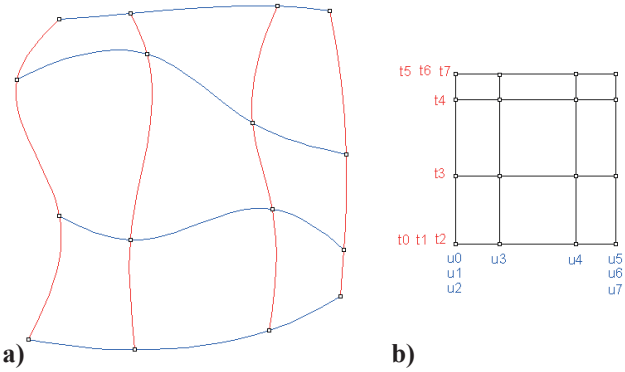
$$\mathbf{S}_{i,j}(t,u) = \sum_{i=0}^{n-4} \sum_{j=0}^{r-4} \mathbf{d}_{i,j} \cdot N_i^3(t) \cdot N_j^3(u) \qquad t_i = \{t_0,..,t_n\}$$

$$u_i = \{u_0,..,u_r\} \qquad (10)$$

where:

t_i, u_i – values of parameters defined for rectangular grid of knots

By contrast to curves, during determining the surface, it is necessary to construct double groups of knots. The first refers to forming the curves; the latter refers to the surface spanning over them. The parameterization for curves and surface can be carried out according to inde-



a) b)

Fig. 5. The grid of knots for making spline surface: a) measured points with the marked curves (red) and with a surface spread in a traverse direction (blue), b) set of initial knots for the curves and surface. (source: own materials)

Rys. 5. Siatka węzłów dla tworzenia powierzchni sklejanych: a) punkty pomierzone z zaznaczonymi krzywymi (czerwone) oraz rozpiętą w kierunku poprzecznym powierzchnią (niebieska), b) zbiór węzłów wyjściowych dla krzywych i powierzchni

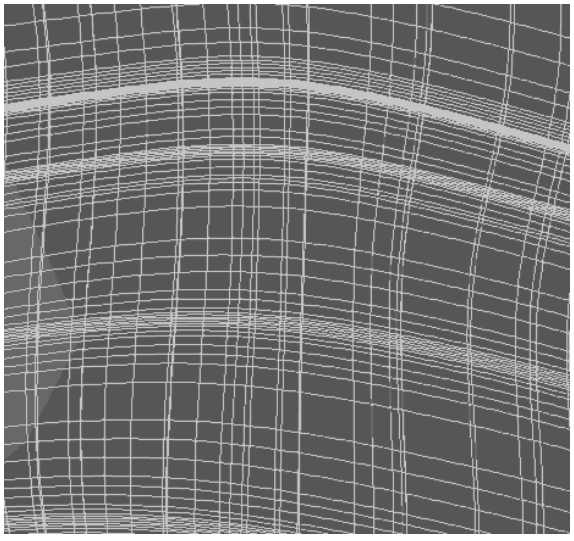


Fig. 6. The section of re-parameterized grids of knots for surface consisting of several dozens of points. (source: own materials)

Rys. 6. Wycinek zreparametryzowanej siatki węzłów dla powierzchni składającej się z kilkudziesięciu punktów

pendent methods. Consequently, both groups make rectangular grid of nodes (Fig. 5.b). This grid, for measurement points chaotically distributed towards one another (Fig. 5.a), it does not consider all the possible parameterizations of each curve, which would consequently cause their deformations. To avoid this, it is used re-parameterization of all curves, meaning such their division, which regards all the cases of parameterization of each curve separately. This leads to greater amount of calculations. The section of re-parameterized grid of knots for the surface consisting of several dozens of points, was shown in figure 6.

5. The example of the model of spline surface

In this point the description of the example spline surface is presented, regarding some aspects discussed in the context of this paper. From the section of the sphere a set of points visible in figure 7.a was taken, then the part of the set ("model points") was selected to create a spline surface (Fig. 7.b). The remaining points („test points") were used to determine the accuracy of the model, by comparison of distances: test points – spline surface. Horizontal spreading of the model was about 13 m. Selected model points were characterized by good regularity in a central part of the surface, remaining on the sides in certain dispersion. Spline surface, formed based on such data, can be characterized by certain waving, the scale of which, will allow to assess the test points. Despite slightly chaotic distribution of model points, it is possible to put on them relatively regularly distributed curves, each of these is spread on relatively regular set of points (Fig. 7.c). The data of this type, allow the conclusion, that it will be possible to create spline surface, possessing relatively small waving. The results of studies comparing distances of test points from the surface, were

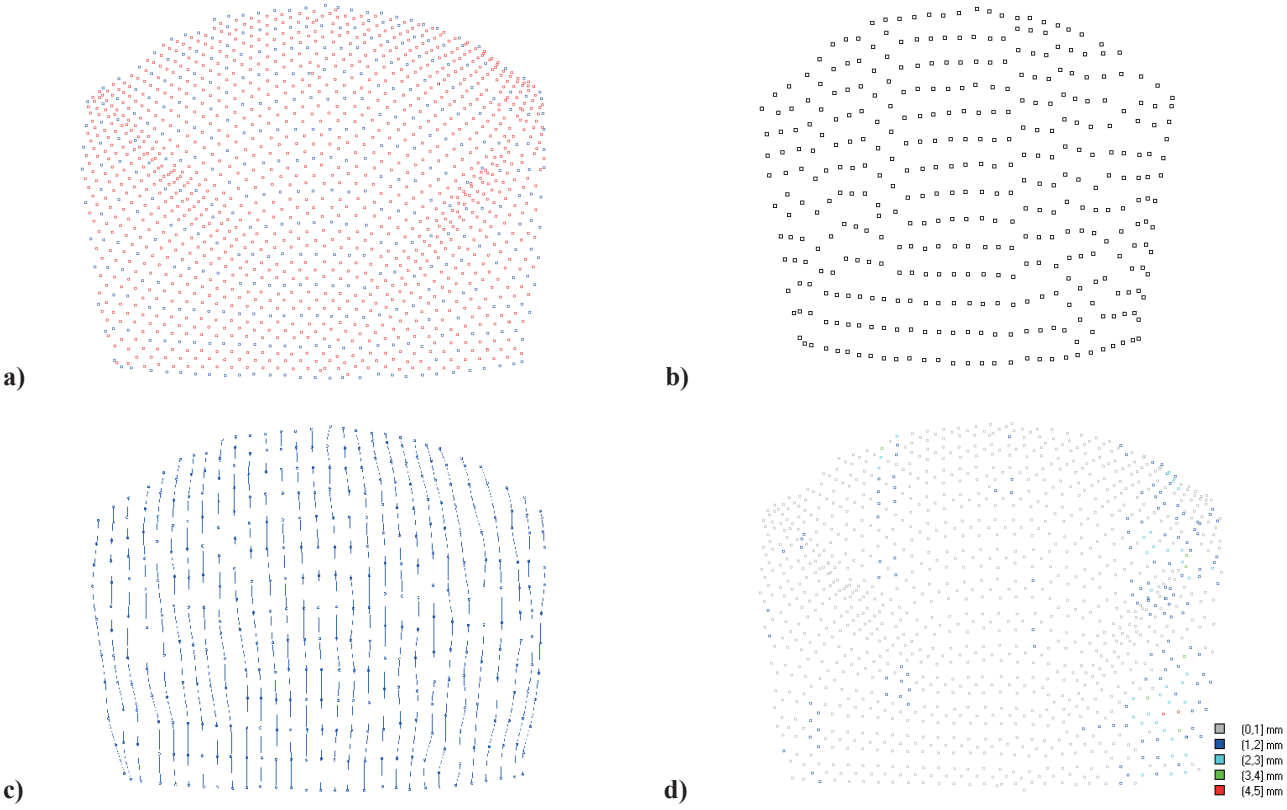


Fig. 7. Formation and accuracy analyses of an example of the spline surface model
Rys. 7. Tworzenie i analizy dokładnościowe przykładowego modelu powierzchni sklejanej

presented in figure 7.d. The areas of good regularity of model points were described with the accuracy below 1mm. In some places the accuracy diminished to about 2mm. The left part of the object was less precisely described. It has less regular distribution of model points. The right part of the surface with more irregular distribution of points, has slightly larger deformations. More numerous is a group of points of deviations reaching 2mm; there is also a certain number of deviations up to 3mm. Sporadically points of deviations as much as 5 mm appeared. Taking into account the size of the object and potential accuracy of reflectorless measurement method (ca. 5–10mm), it can be assessed that, in this case, splines do not bring significant deformations to this model, even despite the dubious regularity of observations in some regions of the surface.

6. Conclusions

Advantages of the description of shape surface with spline functions allow looking at them as a proper tool for building the models of shell structures, especially those of a complicated shell geometry. Some limitations connected with the regularity of the distribution of the measured points and the speed of data processing, are gradually removed through the evolution of the algorithms of the creation and modification of such functions. These algorithms probably will never achieve full efficacy, compensating the lacks in the obtained measurement material, however, already now allows making exact models, impossible to be constructed with other methods. The most advanced works are now connected with making algorithms allowing the processing of the clouds of points coming from laser scanning, which requires more intelligent approximation procedures.

References

Ahlberg J.H., Nilson E.N., Walsh J.L. 1967. The theory of splines and their applications. Academic Press Inc., New York.

Ahn S.J. 2004. Least squares orthogonal distance fitting of curves and surfaces in space. Springer-Verlag, Berlin.

Dan J., Lancheng W. 2006. An algorithm of NURBS surface fitting for reverse engineering. The International Journal of Advanced Manufacturing Technology, Vol. 31(1–2), pp. 92–97.

De Boor C. 1978. A practical guide to splines. Springer-Verlag, New York.

Diercx P.: Curve and surface fitting with splines. Oxford University Press, New York 1995.

Farin G. 2002. Curves and surfaces for computer aided geometric design. Academic Morgan Kaufmann Publishers, San Francisco.

Gocał J. 2010. Geodezja inżynieryjno-przemysłowa, tom III, Wydawnictwa AGH, Kraków.

Haron H., Rehman A., Adi D.I.S., Saba T., Lim S.P. 2012. Parameterization Method On B-Spline Curve. Mathematical Problems in Engineering

Hearn D., Baker M.P. 1987. Computer Graphics, C Version. Prentice Hall, Upper Saddle River, New Jersey.

Kiciak P. 2000. Podstawy modelowania krzywych i powierzchni. WNT, Warszawa.

Kosma Z. 1999. Metody numeryczne dla zastosowań inżynierskich. Politechnika Radomska, Radom.

Lee E. T. Y. 1989. Choosing nodes in parametric curve interpolation. Computer-Aided Design, Vol. 21 No. 6, pp. 363–370.

Lenda G. 2006. Metody tworzenia i modyfikacji funkcji sklejanych dla potrzeb opisu kształtu obiektów obserwowanych punktowo. Geodezja, Vol. 12, No. 2, pp. 277–290, Kraków.

Lim C. G. 1999. A universal parametrization in B-spline curve and surface interpolation. Computer Aided Geometric Design, Vol. 16, No. 5, pp. 408–422.

Safraz M. 2008. Interactive curve modeling with applications to computer graphics, vision and image processing. Springer-Verlag, London.

Shan J., Toth C.K. 2009. Topographic Laser Ranging and Scanning: Principles and Processing. CRC Press.

Tsakiri M., Lichti D., Pfeifer N. 2006. Terrestrial laser scanning for deformation monitoring. 12th FIG Symposium, Baden, May 22–24.

GEOINFORMATICA POLONICA
13: 2014

KONRAD ECKES[1]
eckes@agh.edu.pl

ASSESSMENT OF GEOGRAPHIC SPACE ARRANGEMENT BASED ON THE AMOUNT OF INFORMATION CONTAINED IN THE MAP IMAGE

Keywords:

geographical space, organization and arrangement of space, map image analysis, Shannon's information theory, amount of information in the map image

Abstract

In this paper an attempt was made to determine whether the amount of information, necessary to record a unit area of the geographic space, mapped on the large-scale map, may define the state of the arrangement of this space. The aim of this study was to make a comprehensive assessment of the geographical space in terms of its reasonable development. Intuition suggests that for two areas with similar functions – a picture of orderly and reasonably developed land and a disordered and chaotically developed land – one can record more concisely the image of the first than the latter. The work is an attempt to formulate an indicator that could characterize a state of the general arrangement of the geographical space and could serve as an objective assessment of such arrangement. The proposed statement was put to practical verification.

At the introduction the numerical amount of information was specified, necessary to record unit area of large-scale cadastral maps, containing only the boundaries of the ownership. In the task the concept of the amount of information proposed by C. E. Shannon was used. For a typical map at scale 1:1000, to record the area of one square decimetre (i. e. one hectare in the terrain) 531 bits of information are necessary. For the area of fragmented fields of southeast Poland this number increased to 4980 bits. Basing on previous work of the author, quoted for comparison, the amount of information, necessary to record the identical unit of a typical city in the same coding method, is 23360 bits of information. In this case extensive infrastructure of the city recorded on the source layers has to be included.

Later in the article, the amount of information necessary to record the same unit area of geographical space for the state of extreme disorder (chaos) and the state of the theoretical ideal arrangement was determined. The real determined values were much closer to the ideal in the ordered state than to the chaos. Subsequently, the amount of information was considered necessary for recording sites with different ways of building. Basing on this experience a clear relationship between the amount of information that is necessary to record unit area of geographical space and the degree of rationality of its development could not be established. Thus, the amount of information, required to record of the unit area of large-scale map image, cannot be unequivocal indicator of geographical arrangement and consequently also the image map, because the large-scale map accurately reflects the shape of the geographical space. The growth of the indicator can show the advanced infrastructure of the selected area – on the basis of comparable data.

The indicator of the amount of information can have other wide application – for planning of memory means required to store the geographic space and to determine the transmission time of the recorded data by telecommunication links. The search for a universal indicator of geographical space arrangement, characterizing its organization and rational development – must remain the subject of further research.

[1] AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering.

OCENA UPORZĄDKOWANIA PRZESTRZENI GEOGRAFICZNEJ NA PODSTAWIE ILOŚCI INFORMACJI ZAWARTEJ W OBRAZIE MAPY

Słowa kluczowe:

przestrzeń geograficzna, organizacja i uporządkowanie przestrzeni, analiza obrazu mapy, teoria informacji Shannona, ilość informacji w obrazie mapy

Abstrakt

W pracy została podjęta próba ustalenia, czy ilość informacji, jaka jest konieczna do zapisu jakiegoś jednostkowego wycinka przestrzeni geograficznej, odwzorowanej na mapie wielkoskalowej, może określić stan uporządkowania tej przestrzeni. Celem pracy było dokonanie kompleksowej oceny przestrzeni geograficznej pod kątem jej racjonalnego zagospodarowania. Intuicja wskazuje, że dla dwóch terenów o podobnych funkcjach obraz terenu zagospodarowanego w sposób standardowy i racjonalny można zapisać w sposób bardziej oszczędny niż obraz terenu zagospodarowanego w sposób nieuporządkowany i chaotyczny. Praca jest próbą sformułowania wskaźnika, który mógłby charakteryzować pewien stan generalnego uporządkowania przestrzeni geograficznej, mógłby pełnić rolę obiektywnej oceny takiego uporządkowania. Tak przedstawiona teza została poddana weryfikacji praktycznej.

Na wstępie wyznaczono liczbowe ilości informacji niezbędne do zapisu jednostkowych fragmentów map wielkoskalowych, ewidencyjnych, a więc zawierających tylko granice własności. Do realizacji zadania zastosowano koncepcję ilości informacji zaproponowaną przez C.E. Shannona. Dla typowej mapy w skali 1:1000 do zapisu obszaru jednego decymetra kwadratowego (a więc jednego hektara w terenie) niezbędne jest 531 bitów informacji. Dla obszaru rozdrobnionych działek Polski południowo-wschodniej ta liczba wzrasta do 4980 bitów. Na podstawie wcześniejszych prac autora przytoczono dla porównania, że ilość informacji niezbędna do zapisu identycznego pola jednostkowego typowego miasta, w takim samym sposobie kodowania wynosi 23 360 bitów informacji. W tym przypadku została uwzględniona rozbudowana infrastruktura miasta rozpisana na warstwach źródłowych.

W dalszej części pracy wyznaczono ilości informacji niezbędne do zapisu takiego samego pola jednostkowego przestrzeni geograficznej dla stanu skrajnego nieuporządkowania (chaosu) i stanu teoretycznego uporządkowania idealnego. Realne, wyznaczone wartości okazały się o wiele bliższe stanowi uporządkowania idealnego niż stanowi chaosu. W dalszej kolejności rozpatrywano ilości informacji niezbędne do zapisu terenów o różnych sposobach zabudowy. Na podstawie tych doświadczeń nie udało się ustalić jednoznacznej zależności pomiędzy ilością informacji, niezbędną do zapisu jednostkowego obszaru przestrzeni geograficznej, a stopniem racjonalności zagospodarowania tej przestrzeni. Zatem ilość informacji niezbędna do zapisu jednostkowego obrazu mapy wielkoskalowej nie może być jednoznacznym wskaźnikiem uporządkowania przestrzeni geograficznej i w konsekwencji także obrazu mapy, ponieważ mapa wielkoskalowa oddaje wiernie kształt przestrzeni geograficznej. Wzrost wskaźnika może świadczyć o zaawansowanej infrastrukturze wybranego obszaru – na podstawie porównywalnych danych.

Wskaźnik ilości informacji może mieć inne szerokie zastosowanie – do planowania środków pamięci niezbędnych do zapisu przestrzeni geograficznej oraz do wyznaczania czasu przesyłania tego zapisu przez łącza telekomunikacyjne. Poszukiwanie uniwersalnego wskaźnika uporządkowania przestrzeni geograficznej, wskaźnika charakteryzującego jej organizację i racjonalność zagospodarowania – musi pozostać przedmiotem dalszych badań.

1. Introduction

In the works (Eckes 2002) and (Eckes 2007) attempted to determine the regularities in geographical reality and to determine a formal record of these regularities. Such a record, formed as a set of rules, could become the basis for creating an expert system, which can aid the human factor in verifying large-scale map image. This expert system could provide a test of map image regularities – the homogeneous and objective verification, independent of the personal factors.

At work (Eckes 2007) four thematic groups of expert rules were formed, based on the following relationships:

- relationships contained in the structure of each object images,
- internal relationships between local objects of map image,
- the relationship of objects of the map image to the external physical or geographical factors,
- the relationship of objects of the map image to the externally-shaped patterns.

The above rules allow solving individual tasks.

A more general problem is to make a comprehensive evaluation of geographical space. In the local planning, there are several indicators to assess the geographic space; they are indicators of infrastructure, communication, natural and economic indicators. From the point of view of the effectiveness of the record of this space – we should start from the attempt to formulate an indicator that could characterize a state of the general arrangement of the geographical space and serve as an objective assessment of such an arrangement.

First of all, we should start to determine what we mean by the term "arrangement of geographical space". This term can be defined from the point of view of functional criteria:

- functional completeness – objects of the geographical space fulfil all essential functions of the local community,
- the functional layout of objects necessary for rational action – each object is arranged optimally and easily available to perform its function,
- clear functional division of space – objects are grouped thematically,
- arrangement as compliance with generally accepted standards, developed on rich experiences.

The functional arrangement implies that geographical space, being in the state of order, is more comprehensible, closer to the generally accepted standards or patterns, so its record should be more concise than the record of the disordered one.

Therefore take up the attempt to consider the problem from the side of the amount of information needed to record the map image for areas with similar features, but with a different state of arrangement, with different levels of their organization. Let us try to answer the question, whether the amount of information, that is necessary to record a unit fragment of terrain, can determine the state of its development and arrangement. Intuition suggests that for two areas with similar functions – a picture of the normal and orderly developed land one will be more concise than an image of the disorderly developed land. Let us submit this assumption to practical verification.

2. Analysis of the amount of information required to record of the exemplary map images

The first step in the analysis is to determine the amount of information necessary to record the image of a unit area of the exemplary cadastral map in the scale of 1:1000. Based on statistical calculations on the full section of the map it was determined that on the square 10×10 cm (so in the area of one hectare in reality) there are boundary lines of plots, which consist of eight segments. Each segment is defined by the coordinates of two points. If we assume that to record the coordinates of one point we use twice five-digit numbers – to write such image the set 160 decimal digits is necessary; when they are multiplied by 3.32, this gives us the number of 531.2 binary digits. We can therefore say that for the record of one hectare in the exemplary cadastral map the amount of information equalling 531 bits after rounding off is required.

It should be noted that such a relatively small amount of information per one hectare of land results not only from the larger dimension of rural fields and small unit density of the boundary lines. Another, much

more important factor is the fact that objects are mono-thematic – we determine the amount of information for only one class of objects – for fields, without taking into account a number of other classes as soil quality, buildings and other infrastructure.

Let us now examine other cadastral maps of the south-east Polish territory, with considerably small fields. In this case, in the area of one square decimetre, there are extensive land border lines that make up the average number of 75 segments. After a similar calculation of the decimal digits – one can mark 4980 bits of information as necessary for one hectare of land. In this case we obtained a much greater amount of information for the fixed unit area. But also in this case we are dealing with only one class of objects – the field borders.

Calculation of the amount of information for the unit area, in the case of a larger number of classes of objects is a complex problem: in the map image, in the case of buildings in a city, there is a variety of objects belonging to different classes of infrastructure. These objects can come to different geometric shapes, they can be represented by points, lines and areas.

Calculation of the amount of information falling to the unit area was presented in the article (Eckes 1978). In the task the concept of information proposed by C. E. Shannon (Abramson 1969) was used, in particular the function of entropy of non-memory source. The analysis was carried out on an experimental basis – a part of a large-scale map of the section of Krakow, 640 × 500 mm. Line map (vector map) 1:500, in the area of Kazimierz, presented a typical for many Polish cities downtown area. Experimental calculation of the amount of information for the unit area has been done for the four ways of coding (Eckes 1978). The aim of the experiment was to show the possibility of more efficient coding, approaching to the limit, resulting from the first theorem of Shannon (Abramson 1969).

In the first method of coding independently isolated sections (vectors) was recorded, in the second method takes into account the continuity of the line and the chains of vectors was recorded, in the third and fourth method has made further reduce of code redundancy, in particular, used in the fourth method the incremental coding, allowing to reduce the number of digits required for saving code. The average amount of information per one square centimetre of the map image ranged from 58.4 bits/cm^2 for vector coding – to 40.6 bits/cm^2 for economical incremental coding.

Due to the fact, that in determining the amount of information required for the recording of both units of areas of cadastral maps are taken into account in this experiment recording independent vectors, forming the boundaries of fields – for the sake of comparison let us take the procedure from the article (Eckes 1978), also the value representing the same criteria - the amount of information 58.4 bit/cm^2. This elementary area in the map, in the field corresponds to 0.25 a, so for example for one hectare of exemplary city gives the amount of information of 23360 bits. This number indicates how rapidly the amount of information grows for the area intensively invested, with a high density of details, representing various thematic class of the city infrastructure.

3. Position of the real geographical space between extreme states – chaos and theoretical ideal arrangement

Now, let us consider a possibility to determine what the position of the calculated amount of information for a unit area of land and a city is in the relation to the two extreme states of map image arrangement – ranging from total chaos, to theoretically ideal arrangement of objects of the geographical space.

To determine the level of chaos we shall use the raster model. Assume the same square of the map in scale 1:1000 – 10 × 10 cm, corresponding to one hectare of the terrain. If we assume the resolution of graphical map of 0.1 mm – the raster field will the matrix with dimensions of 1000 × 1000 pixels. If we assume that this will be a binary matrix, which the pixels can take only two values – than is an extreme case, we can get $2^{1\,000\,000}$ different images. Assuming that the appearance of each one of them has the same probability – theoretical maximum amount of information H(S), which is necessary to store one square decimetre (dcm^2) map image (one hectare in terrain), will be:

$$H(S) = \log_2 2^{1\,000\,000} \text{ bits / 1 dcm}^2,$$

after the transformation we get the amount of information H(S) = 1 000 000 bits / 1 dcm^2 map image in a scale 1:1000.

Thus, the limit of maximum chaos one can take the value of one million bits of information per 1 hectare represented on the map in a scale of 1:1000. It should be clear that this value applies to a binary matrix. Taking into account the more numerous domain of the values that can take pixels of the matrix – would extremely increase the amount of information.

To determine the approximate amount of information represented an ideal arrangement – let us assume a fictional, regular division of geographical space on the square fields of side dimension equal to 50 metres. Such fields will form a regular grid, the definition of which requires the coordinates of only one point and the dimension of the mesh. For this 12 decimal digits are enough, which gives 39.8 bits of information.

Then what is the position of information necessary to record the geographical space ranging between extreme chaos and ideal arrangement? Does the map image of geographical space indicate a state closer to the first state or the second of these extreme distant values?

Basing on the above approximate calculation, we can unequivocally state that the image of the geographical space, with numbers expressing quantities of information, 531 bits/1 ha, 4980 bits/1 ha and 23 360 bits/1 ha has in this range, the position much closer to the ideal arrangement than the extreme chaos, even for adopted a binary matrix of the unit image of the map. On the basis of the considerations we are able to determine that the image of the geographical space has a high degree of arrangement, has the state of the advanced organization.

This statement raises the next question: what is the reason that this is happening, and further question – whether the indicator of the amount of information, that represents a unit area of the terrain, can be used as a universal measure of arrangement of the geographical space.

4. Analysis of factors affecting the amount of information written in the form of a large-scale map image

Now, let us consider the problem of the relationship of detail number of geographical space to a number of details recorded as an map image. Geographic space is characterized by extreme richness of details, but only some small important parts of these details are transferred to the map image. The transfer takes into account only the details important from economic and environmental point of view (Eckes 2006).

The first step of the amount reduce is reduce the three-dimensional 3D geographic space on a two-dimensional, by projection of the space on the local reference plane. From the spatial objects we record only their outlines on the ground level.

The second factor which greatly reduces the amount of information transferred on the map is defining the objects. In the process of data acquisition to the large-scale line maps we define point, linear and planar objects forming a wholeness, also object important from the economic and environmental point of view. Thus, in most cases, we take the forest as the area (planar) object (not a single tree), we define the pavement (not a paving block), define the track (not its elements).

The third factor is a simple structure, standard signs used on large-scale line maps – planar objects are represented by the outline of the solid, and the majority of underground cables are represented in the map image as linear signs of the long axis.

In the process of data acquisition a principle of limit density of map image, which makes the picture clear is respected. But in this respect the practice puts higher demands on the amount of details, and even in the case of a traditional analogue maps – maps were made in the version of layers (in Poland four basic layers). Nevertheless, the legibility of the destination of map image is a further limit of the amount of information transferred from the geographical space to map image.

A significant factor limiting the amount of information needed for recording the map image is continuous linear objects and continuous of borders of area objects, in connection with the rectilinear shape of numerous objects or borders of objects of geographical space. The ability of recording of objects using vectors or vector chains is the dominant factor in saving record. Continuous boundaries of planar objects or segments of linear objects are marked only in the form of two end-points in the vector and the end points with the intermediate points – in the case of vector chain.

5. The amount of information contained in the map image and a measure of arrangement of the geographic space

Now take up the attempt to give answer to the last question, may be the amount of information, contained in the map image, an objective indicator to assess the arrangement of the geographical space, to assess the degree of organization of the space, and even the degree of planned development of this area.

In the introduction, this study found that, in accordance with the intuitive reasoning – geographical space in the orderly state of arrangement, should be more comprehensible, closer to the accepted standards or patterns, so the description should be more concise than the description of the disordered one.

We shall verify this thesis on a simple example. Suppose that in the analyzed building square in Kazimierz about 400 people live. Let us consider a similar building square of slum area, located on the outskirts of large cities in Africa or South America. Assume that the identical area of one hectare in 80 small houses also about 400 people live. Building of small houses creates chaos, deprived from functional logic. The area has not any other infrastructure. To record the building one need the set of four vectors for each small house (created by two points). If for the record of every point we need 10 digits for the coordinates – to record such a picture, it is necessary to have the string of 6400 decimal digits, which when multiplied by 3.32 gives us a number of 21 248 bits. Thus, despite the chaos in the buildings in poverty-stricken areas, the amount of information compared to the orderly city buildings (23 360 bytes) – is in this case smaller. The reason for this is building homogeneity in poverty-stricken areas, lack of the developed of infrastructure, the only thematic layer (class objects) are the small houses. On the basis of comparison one can conclude that the arrangement of the geographical space is not tantamount with saving record of information. The amount of information depends not only on the geometry of objects, but also on the number of classes subjected to record, the richness of the infrastructure which includes an exemplary area.

Such a statement can be confirmed with the virtual experience on the cadastral maps quoted at the beginning of this work. If the fields could be divided into parcels, the area is equipped with underground infrastructure and terrain is built – the amount of information, needed to record the identical area, certainly would increase greatly. So, in contrast to the intuitive reasoning – growth of arrangement, planning and organization of objects – it is always associated with an increase in the amount of information needed for recording of unit map image? Is it possible to treat this information as an indicator of the advanced arrangement of geographical space?

Let us consider the area of apartment blocks built in the 1970s and 1980s – extensive multi-stairs blocks built of prefabricated concrete, devoid of any elements which provide aesthetic, typical cuboids, standing in groups of repetitive shapes, provide a little more comfort than the historical buildings of Kazimierz (district of Krakow). Despite the advanced infrastructure, due to the size, less fragmentation of details and simple geometry – it certainly can be saved in a more effective way than the building of the old town.

From this example, two conclusions can be drawn: 30-year-old or 40-year-old buildings give slightly more comfortable living conditions and there is a progress of civilization corresponding to the way of building over centuries. However, the amount of information necessary to record such a building is smaller – due to the size and extreme reduction of the variety of shapes. But this form of building, demonstrating a degree of arrangement and geometric organization of space, does not reflect the progress of civilization. Thus, a smaller amount of information, indicating an increase in geographic space arrangement does not affect the progress of civilization in the development of infrastructure.

6. Conclusion

Is it, therefore, the amount of information, that is necessary to record a single map image, can be an explicit indicator of the arrangement of the geographical space? The above examples show that rather it cannot. The amount of information depends on many factors: the density of the investment, the number of acquisitioned

details, the number of classes of infrastructure and the size of the objects and their geometry. The density of details and the number of classes increases the amount of information, necessary for recording, and size of the objects, geometry and reproducibility can in some cases reduce the amount of information.

Thus, the amount of information required to record the individual unit of map image of large-scale map cannot make an explicit indicator of geographical space arrangement and, consequently, neither the map image, because the large-scale map accurately reflects the shape of geographical space. Growth of the indicator may, to certain extent, reflect the advanced infrastructure of the selected area – on the basis of comparable data.

The indicator of the amount of information can have another wide application – in planning the memory means required to store the geographic space and to determine the transmission time of the recorded data by telecommunication links.

The search for the universal arrangement indicator of geographical space, the indicator characterizing the space organization and rational development – must remain the subject of further research.

References

Abramson N. 1969. Teoria informacji i kodowania. Państwowe Wydawnictwo Naukowe, Poznań (tłumaczenie z języka angielskiego).

Eckes K. 1978. Analiza ilości informacji i wielkości nadmiaru w różnych metodach kodowania mapy numerycznej. PAN – Oddział w Krakowie. Prace Komisji Górniczo-Geodezyjnej, Geodezja 25.

Eckes K. 2002. Logika obrazu mapy jako podstawa do budowy systemu ekspertowego. Geoinformatica Polonica – Prace Komisji Geoinformatyki Polskiej Akademii Umiejętności, nr 4, Kraków.

Eckes K. 2006. Modele i analizy w systemach informacji przestrzennej. AGH Uczelniane Wydawnictwa Naukowo-Dydaktyczne, Seria – Monografie, Kraków.

Eckes K. 2007. Kontrola obrazu mapy wielkoskalowej w systemie eksperckim. Główne problemy współczesnej kartografii 2007; Kartograficzne programy komputerowe – konfrontacja teorii z praktyką. Wyd. Uniwersytet Wrocławski, Instytut Geografii i Rozwoju Regionalnego, Zakład Kartografii, Wrocław.

GEOINFORMATICA POLONICA
13: 2014

IZABELA KARSZNIA[1]
i.karsznia@uw.edu.pl

AN ATTEMPT OF FORMALIZING THE SELECTION PARAMETERS FOR SETTLEMENTS GENERALIZATION IN SMALL-SCALES

Key words:

Spatial data generalization, settlements selection, generalization parameters

Abstract

The paper covers one of the most important problems concerning context-sensitive settlement selection for the purpose of the small-scale maps. So far, no formal parameters for small-scale settlements generalization have been specified, hence the problem seems to be an important and innovative challenge. It is also crucial from the practical point of view as it is necessary to develop appropriate generalization algorithms for the purpose of the General Geographic Objects Database generalization which is the essential Spatial Data Infrastructure component in Poland.

The author proposes and verifies quantitative generalization parameters for the purpose of the settlement selection process in small-scale maps. The selection of settlements was carried out in two research areas – in Lower Silesia and Łódź Province. Based on the conducted analysis appropriate contextual-sensitive settlements selection parameters have been defined. Particular effort has been made to develop a methodology of quantitative settlements selection which would be useful in the automation processes and that would make it possible to keep specifics of generalized objects unchanged.

PRÓBA FORMALIZACJI DOBORU PARAMETRÓW GENERALIZACJI MIEJSCOWOŚCI DLA OPRACOWAŃ W SKALACH PRZEGLĄDOWYCH

Słowa kluczowe:

Generalizacja danych przestrzennych, wybór miejscowości, parametry generalizacji

Abstrakt

Przedmiotem artykułu jest kontekstowy dobór parametrów selekcji miejscowości dla opracowań w skalach przeglądowych. Jest to zagadnienie ważne oraz innowacyjne, gdyż do tej pory nie ustalono i nie zweryfikowano formalnych kryteriów selekcji miejscowości w skalach małych. Poruszany problem ma także duże znaczenie praktyczne ze względu na rozwój infrastruktury informacji przestrzennej w Polsce, w szczególności pod względem opracowania procedur generalizacji Bazy Danych Obiektów Ogólnogeograficznych.

[1] Department of Geoinformatics, Cartography and Remote Sensing, Faculty of Geography and Regional Studies, Warsaw University

Autorka przedstawia koncepcję regionalnego zróżnicowania parametrów selekcji miejscowości w ramach poszczególnych powiatów na dwóch obszarach badawczych: w województwie dolnośląskim i łódzkim. W proponowanej metodyce selekcji miejscowości uwzględniono zarówno charakterystyki atrybutowe, jak i przestrzenne sieci osadniczej, co umożliwiło poprawne odzwierciedlenie struktury wielkości i ważności miejscowości, a także zachowanie specyfiki oraz relacji przestrzennych między nimi. Opracowane procedury generalizacji mogą zostać wykorzystane do uogólnienia sieci osadniczej w Bazie Danych Obiektów Ogólnogeograficznych.

1. Introduction

The reduction of the map scale inevitably leads to the emergency of spatial conflicts between geographic objects. In order to maintain a visual clarity, it is necessary to adjust the level of detail of the map content to the scale and purpose of the presentation (H. Yan, R. Weibel 2008). That comprises the operations of selecting proper information, its simplifying, aggregating and setting out symbolization. These actions constitute the process of generalization, which aims at improving the communication skills of the map or visualization, and also highlighting the characteristics of a presented phenomenon (J. Tyner 2010). The dynamic development of geographic information systems, a need for effective processing and visualizing large amounts of spatial data as well as spatial data infrastructure advancement makes up a direct cause of seeking solutions for automated databases generalization (C.B. Jones, J.M. Ware 2005, G. Touya, J.F. Girres 2013).

In this paper we have analyzed one of the most important problems of effective and proper generalization of spatial data, namely, context-sensitive settlements' selection for the purpose of the small-scale maps. After a wave of research related to the generalization of buildings and the built-up areas of large scale levels (R. Weibel and L. Harrie 2007, J. Stoter et al. 2009; V. Van Altena et al. 2013), it is undoubtedly an up-to-date and innovative issue. Undertaking this research problem is also justified by the fact that no formal parameters for small-scale settlements generalization have been specified so far. It is necessary to develop appropriate generalization algorithms for the purpose of generalization of the General Geographic Objects Database, which is the key element of the Spatial Data Infrastructure in Poland (D. Gotlib et al., 2007; K. Chałka et al. 2011). This study is a further step towards the development and implementation of required generalization procedures.

2. The goal and the scope of the research

The aim of this study is to propose and verify the quantitative generalization parameters for the small-scale maps. Selection of the settlements was carried out in two research areas – in Lower Silesia and Łódź Provinces. Both regions are quite diverse in terms of the settlement type, while typical for the central and south-western parts of the country. The wide variations in the settlements structures size in the analyzed regions unable it to use a single criterion for the settlements selection. Therefore, it is necessary to divide the province area into smaller administrative units (districts). Then, it is necessary to adjust the settlements' selection parameters accordingly to the local settlements' characteristics like settlements' density and population density. In order to conduct such local differentiation analysis the relationship between settlements' density and population density had to be conducted. The proposed concept of regional differentiation of generalization parameters can be implemented in the system supporting the generalization of spatial data in order to automate the quantitative generalization of the settlements.

3. State-of-the-art in the small-scale maps generalization

Selection of the settlements that will be displayed on the map at a smaller scale is one of the first steps taken into account in the process of spatial data generalization (F.T. Topfer and W. Pilewizer 1966; N. Kadmon, 1972; G.E. Langran and T.K. Poiker 1986; D.E. Rich-

ardson and J.C. Muller 1991). Intuitively we feel that larger cities (measured primarily in the population value) should take precedence over smaller ones in terms of their appearance on the target map. In practice, the cities selected to be represented on the target map are not always the largest, but the most important or characteristic ones. For instance, when a large city is located in close proximity to other large one, it may not be shown on the target map. At the same time a smaller city, located in isolation from other cities may be kept, by virtue of its "relative importance" (M. Van Kreveld et al. 1997). Consequently, the primary issue is the optimal choice of the settlements, considering not only their size but also their importance and characteristics (I. Karsznia 2011abc).

Several quantitative generalization criteria of the settlements selection have been proposed so far. (K. Podlacha 1966, L. Ratajski 1973, M. and W. Grygorenko Baranowski 1974; M. Sirko 1988, 1991, D.M. Flewelling and M.J. Egenhofer 1993; M. van Kreveld et al. 1997; I. Karsznia 2010, I. Karsznia 2011abc).

The generalization algorithms of point objects proposed until now have been described by Li Z. (2007). The author has divided them into two groups: the algorithms for so-called selective omission of point objects and the algorithms for the simplification of the structure of a set of point objects.

The first group of algorithms is intended primarily for the generalization of the settlements. When reducing the map scale, not all point objects (settlements or cities) are shown as a result of the less important ones should be omitted. Among the existing selection algorithms, Z. Li distinguishes the following ones: a settlement-spacing ratio model, gravity modeling, distribution-coefficient control model, set segmentation and quadrat-reduction models (Li 2007). In the last two solutions there is too much manual intervention required thus they are not efficient in the automatic settlements' generalization process.

The problem of effective settlement selection for interactive visualization has also been addressed by M. van Kreveld et al. (1997). The authors propose a settlement selection model called circle growth with its two modifications. In the article the opportunities and constraints of the model were also described. The subject of the research is not to rank the cities (the authors assume that each city has already been classified in terms of its importance), but the cities selection process itself. The goal is to take into account and solve the spatial conflicts, which aroused during the generalization process, as well as to keep and even highlight the geographical context of the generalized objects.

The second group of generalization algorithms described by Z. Li (2007) was designed to simplify the structure of the set of point objects based on a set of parameters describing it. In the implementation process of these algorithms both attribute and spatial characteristics of the set of point objects have been taken into consideration. These include: the quantity, the weight of each object point, their proximity and the relationships among them, the spatial distribution of objects as well as their density. This group of algorithms includes algorithms, which take into account geometric objects characteristics (T. Ai, Y. Liu 2002) and algorithms which take into account both geometric and thematic information, such as statistics, thematic or topological (T.E. Samsonov, A.M. Krivosheina 2012).

However, due to the limitations of poor structure of existing spatial databases, not always can we use the existing generalization algorithms entirely (I. Karsznia 2010). The alternative is to enrich the structure of the source database with additional important spatial and attribute characteristics of objects which enable the correct selection of the cities. These tasks are referred to as the data enrichment, and the process of analyzing groups of objects and relationships among them and thus map context – is often called the structure recognition process (S. Steiniger, R. Weibel 2007). The term adaptive generalization means the process of making decisions which involves both the selection of a generalization operator to a particular group of objects and corresponding parameter values (constraints) to the specific graphic situation (M. Neun 2007).

4. The regional differentiation of the settlements selection parameters concept

As proved in the previous studies (I. Karsznia 2010, 2011abc) the introduction of a one quantitative set-

tlements' selection parameter for the province region is not a good solution because it does not provide the variation in the structure size and density of the settlements correctly. In order to allow the contextual settlements' selection and its proper visualization it has been proposed an approach of regional differentiation of the quantitative generalization parameters within individual districts.

4.1. Research assumptions

Based on the literature and the ongoing research in this area it was possible to formulate four basic assumptions. The leading criterion for the settlements selection is their size, expressed in the population value (K. Podlacha 1966, J. Ostrowski 1970). Moreover, due to the need to maintain the correct settlements size structure it has been assumed that the differences in density of the settlements in a given region (province) should reflect the diversity of population density. Regarding that, the density index of the settlements, expressed as the amount of settlements falling on 1 dm^2 for urban areas on Polish territory should be in the range of 50 to 150 settlements. The differentiation of the quantitative generalization parameters of settlements was carried out in the districts. The urban districts were excluded from the analysis; due to their specific nature (the existence of one large city precludes density analysis). Subsequent analyses were focused on rural districts.

4.2. Analysis of the settlements density and its dependence on the density of population

In order to make the regional differentiation it was essential to investigate the detailed relationship analysis between the density, which was selected for various size criteria of settlements, and the density of population. To determine which one of the assumed density of settlements (over 100, 200, 300 or 400 inhabitants) better captures the density of population in the various districts the graphs of density of settlements and population density were constructed and than compared with each other in particular districts of two provinces. The charts showing settlements density and population density in the districts of Lower Silesia and Łódź are shown in Figures 1 and 2 (I. Karsznia 2010; I. Karsznia 2013).

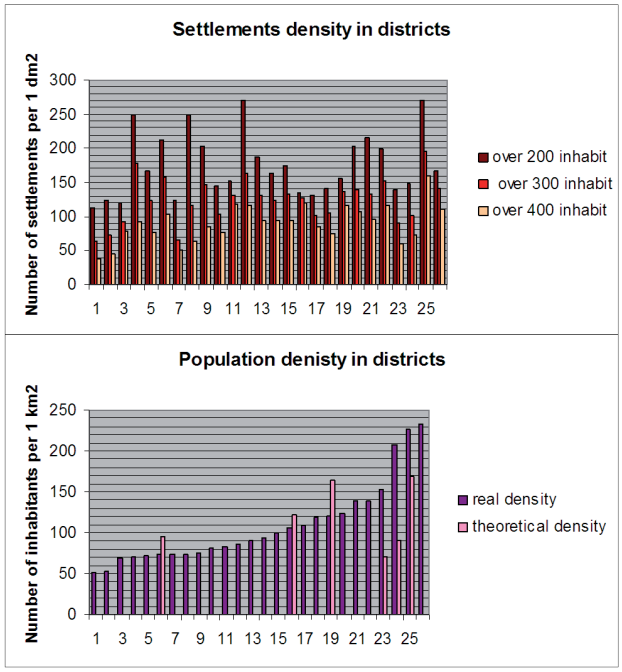


Fig. 1. A chart of the settlements density and the density of population relationship in Lower Silesia
Rys. 1. Wykres zależności gęstości miejscowości oraz gęstości zaludnienia w poszczególnych powiatach w województwie dolnośląskim

In addition to the urban districts having its own characteristics that were excluded from the analysis, in both provinces there are other two districts groups which would be given a more thorough evaluation. These are districts including large cities what results in a situation that a population density in a district is relatively high, whilst settlements density is quite low (such as Lubin District in Lower Silesia or Radomsko District in Łódź Province). Furthermore, in districts from which a large city has been excluded and based on that an urban district has been created the population density is quite low (Legnica District in Lower Silesia, or Pitroków in Łódź Province). A large city has been defined as a city with the population more than 50 000 inhabitants, which is slight-

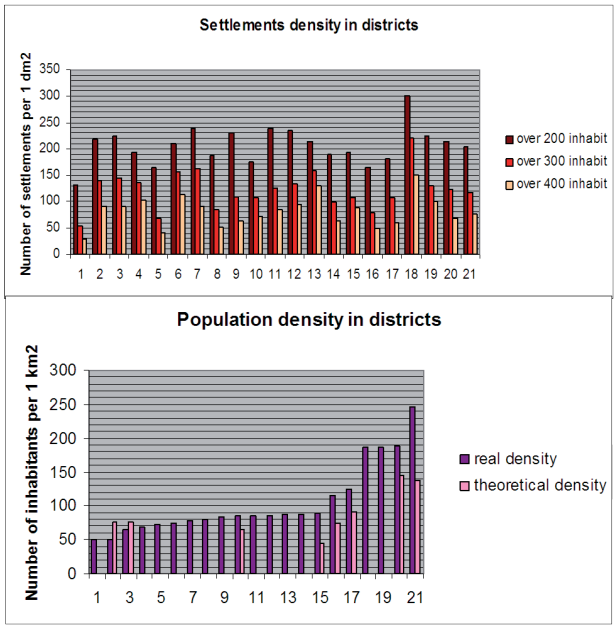


Fig. 2. A chart of the settlements density and the density of population relationship in Łódź Province

Rys. 2. Wykres zależności gęstości miejscowości oraz gęstości zaludnienia w poszczególnych powiatach w województwie łódzkim

ly more than half of the average number of people in the analyzed districts. The districts of "overvalued" and "undervalued" population have been denoted accordingly as P_{zw} and P_{zn} [2].

To ensure comparability across all districts in particular provinces, it was necessary to conduct some kind of "normalization" of all the districts. For this purpose the concept of "average" (typical) city in the district, marked with $P_{śr}$ has been proposed. Subsequently, the theoretical average population of the typical city has been formulated as $LP_{śr}$.

$$LP_{Śr} = \frac{\sum\left(LP - \left(LP_{ZW} + LP_{ZN}\right)\right)}{\sum\left(P - \left(LP_{ZW} + LP_{ZN}\right)\right)}$$

In order to obtain comparability among all the districts within provinces the population of the average city was added to the undervalued districts and then the population density was calculated. However, in districts with an over valuated population density the largest city was replaced with the "average" city and hence, the theoretical density of population has been calculated. The comparison between the settlements density and population density (the real and the theoretical one for specific districts – see Figures 1 and 2) allowed us to make a final decision which of the proposed settlements density (more than 200, 300 or more than 400 inhabitants) better reflects (is "closer" in terms of its value) the population density in each of the districts. The conducted analysis has led to proposing a regional differentiation of the settlements selection parameters in particular districts. As a result in Lower Silesia Province the settlements density with a population over 400 inhabitants proved to be optimal in the seventeen districts and the

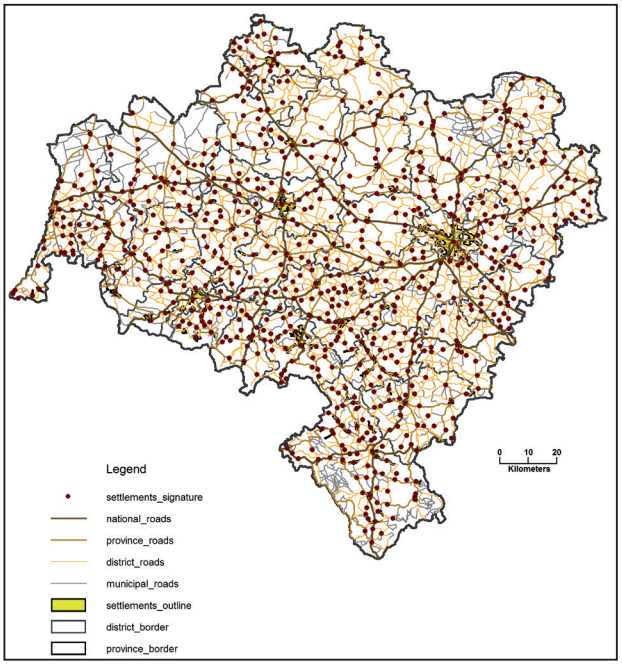


Fig. 3. The results of the settlement selection process based on proposed regional differentiation of the settlements within districts in Lower Silesia Province

Rys. 3. Wybór miejscowości w poszczególnych powiatach w województwie dolnośląskim dokonany zgodnie z koncepcją regionalnego zróżnicowania parametrów selekcji miejscowości

[2] A *P* denotes a district, *LP* denotes district population

population density over 300 inhabitants in nine districts. Similarly in Łódź Province the settlements density with a population over 400 inhabitants proved to be optimal in the twelve districts and the population density over 300 inhabitants in nine districts. Figures 3 and 4 present the results of regional differentiation of the settlements selection parameters within districts respectively for Lower Silesia and Łódź Province.

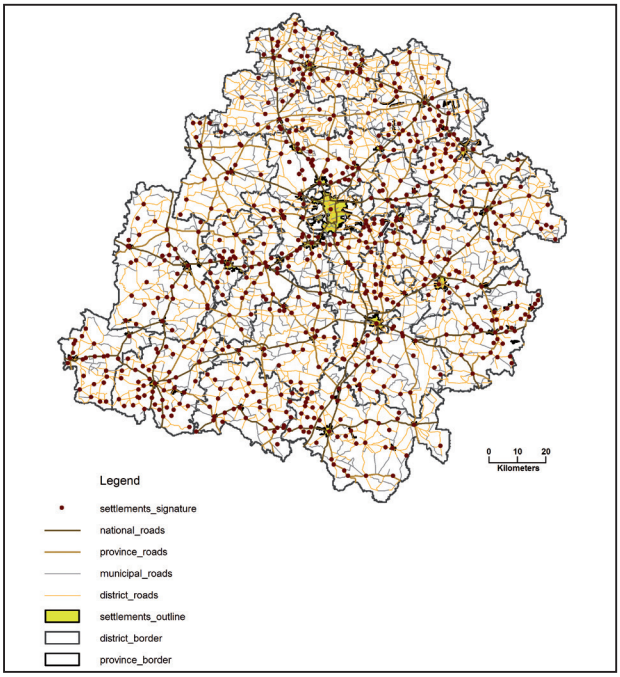


Fig. 4. The results of the settlements selection process based on proposed regional differentiation of the settlements within districts in Łódź Province

Rys. 4. Wybór miejscowości w poszczególnych powiatach w województwie łódzkim, dokonany zgodnie z koncepcją regionalnego zróżnicowania parametrów selekcji miejscowości

5. Conclusions

Map generalization is responsible for deriving from a source database a target map or database at a reduced scale, whose contents and detail have been reduced to preserve the structural characteristics of the source data. The research described within this article falls squarely within the scope of the current research context of automated solutions for quantitative generalization.

This paper provides an attempt of formulating principles of quantitative settlements generalization and relating them to the reference data contained in the Polish General Geographic Database.

In the proposed methodology of quantitative settlements' selection from the source scale of 1:250 000 to the target scale of 1:500 000 some important factors were taken into account. First of all the reference character of the database, which enables them to provide a single reference system for collecting additional spatial data was considered.

Another important factor is a database destination. It can be used in the administration and management at the central level but it can also be used for analysis and decision making processes of strategic importance. It can also be used as a source and background data for elaborating thematic, administrative, economic and other maps.

Particular effort has been also made to develop a methodology of quantitative settlements selection which would be useful in automated processes with keeping specifics of generalized objects unchanged. The obtained results look promising, however they require further investigations in terms of verification of the proposed methodology on various test areas, including for instance mountain or coastal areas, as well as its further integration with other thematic map layers generalization, like for example road network.

References

Ai T., Liu Y. 2003. Analysis and simplification of point cluster based on Delaunay triangulation model. [In:] Advances in spatial analysis and decision making, Li Z., Zhou Q., Kainz W., (eds), Londyn, Taylor and Francis, pp. 9–19.

Baranowski M., Grygorenko W. 1974. Próba obiektywnego doboru osiedli na mapie z zastosowaniem maszyny cyfrowej. Polski Przegląd Kartograficzny, Vol. 6, No. 4, pp. 149–155.

Chałka K., Olszewski R., Zieliński J. 2011. Bazy Danych Obiektów Topograficznych i Ogólnogeograficznych – zakres merytoryczny i techniczny opracowywanego projektu rozporządzenia MSWIA, Roczniki Geomatyka, Vol. 9, pp. 98–102.

Flewelling D.M., Egenhofer M.J. 1993. Formalizing importance: parameters for settlement selection from a geographic database, Materiały konferencyjne Auto-Carto XI, Minneapolis.

Gotlib D., Iwaniak A., Olszewski R. 2007. Harmonizacja baz danych referencyjnych – niezbędny krok ku budowie krajowej infrastruktury danych przestrzennych. Roczniki Geomatyka, vol. 5, pp. 35–38.

Harrie L., Weibel R. 2007. Modelling the Overall Process of Generalisation. [In:] Mackaness, W.A., Ruas, A. and Sarjakoski, L.T. (eds.). Generalisation of Geographic Information: Cartographic Modelling and Applications. Elsevier Science, pp. 67–87.

Jones C.B., Ware J.M. 2005. Map generalization in the web age. International Journal of Geographical Information Science 19 (8–9), pp. 859–870.

Kadmon N. 1972. Automated selection of settlements in map generation. The Cartographic Journal, Vol. 9, No. 2, pp. 93–98.

Karsznia I. 2010. Podstawy metodyczne automatyzacji generalizacji wybranych elementów Bazy Danych Ogólnogeograficznych. Praca doktorska, Uniwersytet Warszawski.

Karsznia I. 2011a. Z problematyki generalizacji osadnictwa i sieci dróg na mapach przeglądowych – metodyka i narzędzia generalizacji. Cz I. Polski Przegląd Kartograficzny, Vol. 43, No. 1, pp. 5–26.

Karsznia I. 2011b. Z problematyki generalizacji osadnictwa i sieci dróg na mapach przeglądowych – metodyka i narzędzia generalizacji. Cz II. Polski Przegląd Kartograficzny, Vol. 43, No. 2, pp. 117–129.

Karsznia I. 2011c. On automatic generalization of General Geographic Database in the applications of geographic information systems. Czasopismo Geograficzne, 82(1–2), pp. 189–205.

Karsznia I. 2013. Selected aspects of settlements generalization in the General Geographic Database in Poland. Journal of Geodesy and Cartography, 39(3), pp. 129–137.

Langran C.N., Poiker T.K. 1986. Integration of name selection and name placemen. Proceedings of 2nd International Symposium on Spatial Data Handling, pp. 50–64.

Li Z. 2007. Algorithmic Foundation of Multi-Scale Spatial Representation. CRC Press.

Neun M. 2007. Data enrichment for adaptive map generalization using web services. Praca doktorska, Uniwersytet w Zürichu.

Ostrowski J. 1970. Analiza obrazu osiedli na Mapie Świata 1:2,5 M. „Polski Przegląd Kartograficzny", Vol. 2, No 1, pp. 1–13.

Podlacha K. 1966. Generalizacja osiedli wiejskich. Prace Instytutu Geodezji i Kartografii, Vol. 13, No. 3, pp. 4–84.

Ratajski L. 1973. Rozważania generalizacji kartograficznej. Polski Przegl. Kartogr., Vol. 5, No. 2, pp. 49–54 (Part I); No 3, pp. 103–110 (Part II).

Richardson D.E., Muller J-C. 1991. Rule selection for small-scale map generalization. [In:] Map generalization: making rules for knowledge representation, ed. B. Buttenfield, R. McMaster, London, Longman, pp. 136–149.

Topfer F., Pillewizer W. 1966. The principles of selection: a means of cartographic Generalization. The Cartographic Journal, Vol. 3, No. 1, pp. 10–16.

Touya G., Girres J-F. 2013. ScaleMaster 2.0: a ScaleMaster extension to monitor automatic mulit-scales generalizations. Cartography and Geographic Information Science, Vol. 40, No. 3, pp. 192–200.

Tyner J. 2010. Principles of map design. The Guilford Press, New York, London.

Samsonov T.E., Krivosheina A.M. 2012, Joint generalization of city points and road network for small-scale mapping. [in:] Materiały konferencyjne GIScience, siódmej Międzynarodowej Konferencji na temat systemów informacji geograficznej, Ohio.

Sirko M. 1988. Teoretyczne i metodyczne aspekty obiektywizacji doboru osiedli na mapach. Rozprawy Wydziału Biologii i Nauk o Ziemi Uniwersytetu Marii Curie-Skłodowskiej, Rozprawy Habilitacyjne nr 32, Lublin.

Sirko M. 1991. Podstawy metodyczne generalizacji osiedli na mapach. Materiały VII Szkoły Kartograficznej. Generalizacja kartograficzna. Polanica-Zdrój, Wydanie Uniwersytetu Wrocławskiego, pp. 33–36.

Steiniger S., Weibel R. 2007. Relations among map objects in cartographic generalization. Cartography and Geogr. Inform. Science, Vol. 34, No. 3, pp. 175–197.

Stoter J., Burghardt D., Duchêne C., Baella B., Bakker N., Blok C., Pla M., Regnauld N., Touya G., Schmidt S. 2009. Methodology for evaluating automated map generalization in commercial software. Computers, Environment and Urban Systems, 33(5), pp. 311–324.

Van Altena V., Post M., Nijhuis R., Stoter J. 2013. Generalisation of 1:50k roads centerlines from 10k polygons in an automated workflow. 16th ICA Workshop on Generalisation and Map Production – Dresden, Germany. http://generalisation.icaci.org/images/files/workshop/workshop2013/genemappro2013_submssion_18.pdf

Van Kreveld M., Van Oostrum R., Snoeyink J. 1997. Efficient settlement selection for interactive display. Materiały konferencyjne Auto-Carto XIII, Seattle. http://mapcontext.com/autocarto/proceedings/auto-carto-13/pdf/efficient-settlement-selection-for-interactive-display.pdf

Ware J.M., Jones C.B. 2005. Map generalization in the web age. International Journal of Geographical Information Science, 17(8), pp. 743–769.

Yan H., Weibel R. 2008. An algorithm for point cluster generalization based on the Voronoi diagram. Computers and Geosciences, No. 34, pp. 939–954.

GEOINFORMATICA POLONICA
13: 2014

KONRAD NERING[1]
knering@mech.pk.edu.pl

SHORT-TERM FLOOD PREDICTION SYSTEM

Key words:

hydrologic model, geographic information system, hydrologic forecast, flood warning system

Abstract

This paper describes a fully functional short-term flood prediction system. Its effect has been tested on watershed of Lubieńka river in Małopolska. To use this system it must have a data set also described in this paper. A modification of the system to adopt for predicting flash floods was described. Full operation of the system is shown on example of real flood on Lubieńka river in June 2011.

SYSTEM KRÓTKOTERMINOWEJ PROGNOZY HYDROLOGICZNEJ

Słowa kluczowe:

model hydrologiczny, systemy informacji geograficznej, prognoza hydrologiczna, system ostrzegania powodziowego

Abstrakt

W niniejszym artykule opisano w pełni działający system krótkoterminowej prognozy hydrologicznej. Jego działanie zostało przetestowane na zlewni potoku Lubieńka w województwie małopolskim. Aby uruchomić opisywany system, należy dysponować szeregiem danych, których proces przygotowania został zaznaczony w artykule. Opisano także autorską modyfikację zastosowanego modelu hydrologicznego polegającą na dostosowaniu jego działania do tworzenia prognoz hydrologicznych. Pełne działanie systemu przedstawiono na przykładzie wezbrania na potoku Lubieńka z czerwca 2011 roku.

1. Introduction

One of the most dangerous phenomena in nature is flash flood. This is a rapid phenomena that cause overflow or accumulation of an expanse of water that submerges low-lying areas. This phenomenon cannot be avoided. However, damages caused by flood can be minimized.

There are many methods of flood control for minimizing flood damage i.e. building dams, weirs, levees, etc. or creating warning systems with prediction of floods. Building any of these hydro-technical facilities is very expensive and sometimes impossible to locate on small watersheds. For this kind of watershed (area less than 50 km^2) a better solution is to create a short term

[1] Tadeusz Kościuszko Krakow University of Technology, Faculty of Mechanical Engineering, Division of Fluid Mechanics.

flood prediction system to predicting flash floods. It also can minimize property damage (i.e. cars).

Creating a flood prediction system needs an efficient hydrological model [3]. There are many types of hydrological models, but the best for flood predicting in small watersheds is a deterministic type [15].

In this paper a short-term flood prediction system was presented. This system contain hydrological model adapted to local conditions of selected small watershed. Created system generates a short-term hydrological forecast for the area of watershed.

A small watershed was located in małopolskie voivodeship near Myślenice. It was a watershed of Lubieńka river (Fig. 1). A model used to creating hydrological forecast was CASC2D with modifications.

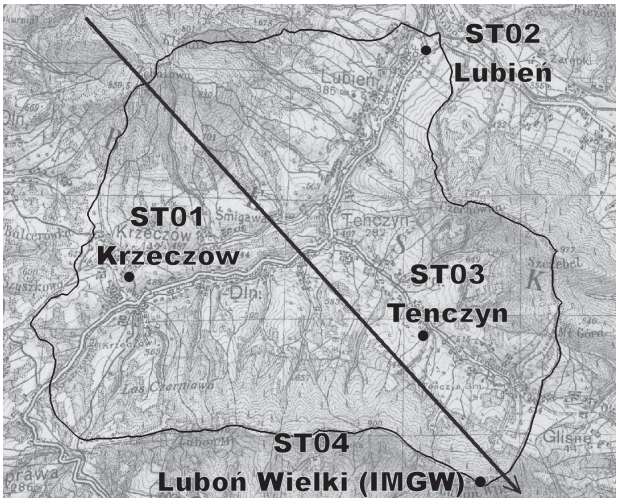


Fig. 1. Lubienka River Watershed map
Rys. 1. Obszar zlewni testowej potoku Lubieńka

2. Digital Elevation Model

The CASC2D (CASCade 2 Dimensional SEDiment) is a fully-unsteady, physically-based, distributed-parameter, raster (square-grid), two-dimensional, infiltration-excess (Hortonian) hydrologic model for simulating the hydrologic response of a watersheds subject to an input rainfall field [2, 6]. This model bases on set of data that consist of digital elevation model, watershed boundary, channel map, land use, and spatially distributed precipitation map.

First set of data contain elevation data. To create this set two public digital elevation models (DEM) was compared. SRTM data [8,9] and ASTER data [7]. Resolution of these DEM data is 1 arc-second (about 30 m raster). After comparing SRTM and ASTER data with topographic map (1:10 000) it can be observed that ASTER set is more reliable for selected watershed. Fig. 2 shows comparison of elevation profile for NW-SE line presented on Fig. 1.

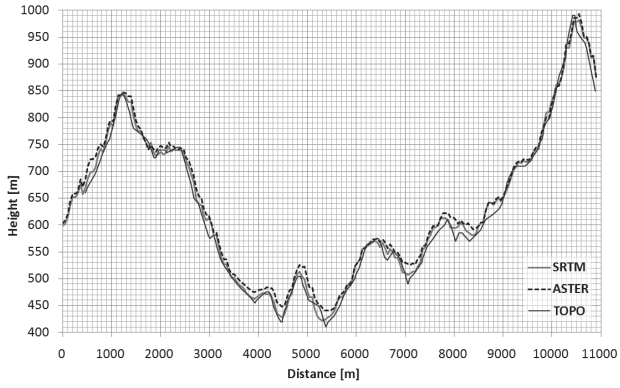


Fig. 2. Comparision of elevation profile generated from different DEM
Rys. 2. Profile wysokościowe dla różnych danych NMT

SRTM data model is significantly different from the elevation of the topographic map – up to 40 m. ASTER data was selected, because they are up to date and give more reliable information about terrain. But this set is not perfect and their limitation should be taken into account, especially in the case of hydrologic simulation based on this data set.

Using DEM data can be determined area of the watershed automatically. It was done by using one of the GRASS GIS module package called *r.watershed* [12]. Depending on DEM sets different catchment area was determined. The algorithm for determining watershed boundary calculate the maximum height from which the water will flow down the valley by gravity force [10]. Thus any errors in the DEM data imply errors in determining watershed boundary. Therefore, it is important to calculate watershed area with one selected DEM to avoid errors

appearing later in a hydrological model. These errors are small in terms of accuracy of the calculations however they can lead to instability of the numerical model.

3. Channel Network

Next set of data used in CASC2D model is the river network in the watershed. This data set can be easily prepared with the digital elevation model. There is key advantage creating channel network using DEM. Thus the designated channels will certainly be at the lowest point of the valley section. To determine channel network an algorithm implemented in GRASS GIS was used. The module *r.stream.extract* calculate streams position by finding the lowest point of the valley section. Fig. 3. shows a comparison between the algorithm of *r.stream.extract* module and exact streams position from the topographic map. There are some differences related to the imperfection of the ASTER DEM. However, they do not affect significantly the results of the hydrological simulation. Some streams marked on topographic map were not included due to the fact that they are temporary.

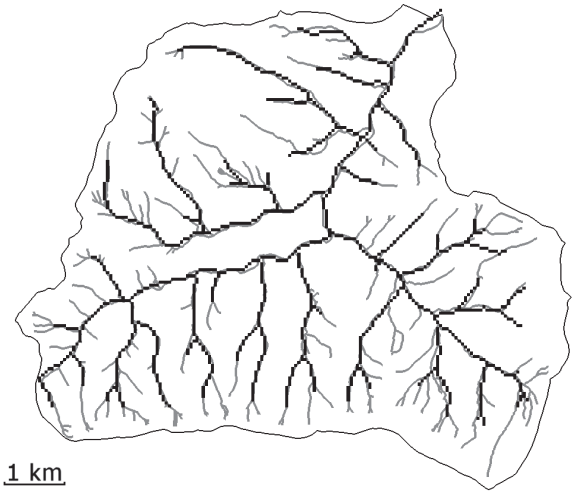


Fig. 3. Stream network based on ASTER DEM (black) and stream network based on topographic map (grey)

Rys. 3. Sieć rzeczna wyznaczona na podstawie danych wysokościowych ASTER (czarny) w porównaniu do danych topograficznych (szary)

4. Land use

An important element of the set of input data for hydrological model CASC2D is a map of the land use. There are many forms of land use affecting water balance in the watershed area, i.e. deciduous forests, coniferous forests, meadows, pastures, buildings, urban infrastructure, etc. It can be assumed that these data are fixed, but only in short periods (i.e. couple of years). To simplify the set of data three basic types of land use was selected: forests, meadows, urban areas. This simplification was made because the area of selected watershed is dominated by mixed forests, meadows and pastures, and minor urban areas.

Land use spatial data can be achieved in at least two ways. The first is to review the topographical map. This operations can be performed with GRASS GIS using library of modules responsible for the analysis of the raster maps.

Spatial data from topographic maps are available in large resolutions (up to 1 meter), but have a significant disadvantage: they are not always up to date. To get the most recent data about land use the second method

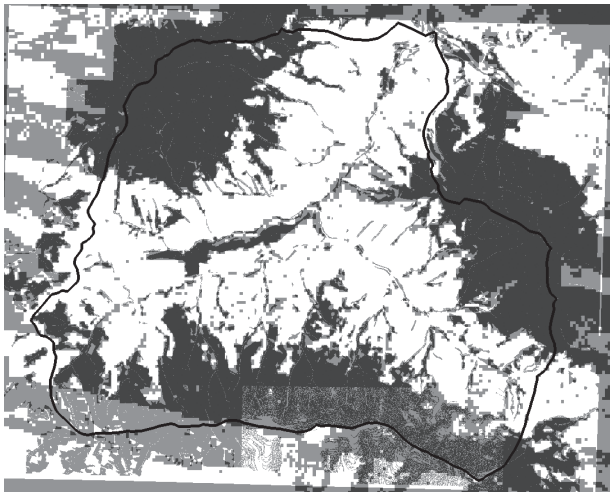


Fig. 4. Forest map from LANDSAT-7 (dark-grey) and from topographic map (light-grey)

Rys. 4. Pokrywa leśna z danych satelity LANDSAT-7 (ciemnoszary) i danych z mapy topograficznej (jasnoszary)

can be used by extracting data from satellite images of Landsat-7 (not older than one month). Unfortunately the resolution of these raster data is approximately 60 meters. However, for the hydrological simulation in CASC2D it is good enough. Fig. 4 presents a comparison of Landsat data and topographic data relating to forest cover in the examined watershed (July 2010).

5. Soils

The last set of fixed spatial data is the spatial distribution of soils in the watershed area. These data are important because of the description of water flow when infiltration occurs. Soils map (Fig. 5) was developed on the base of Mapa Gleb Polski (1:300 000) [5]. The accuracy in this case (using CASC2D model) is sufficient. Map of soils spatial data was prepared using GRASS GIS packages.

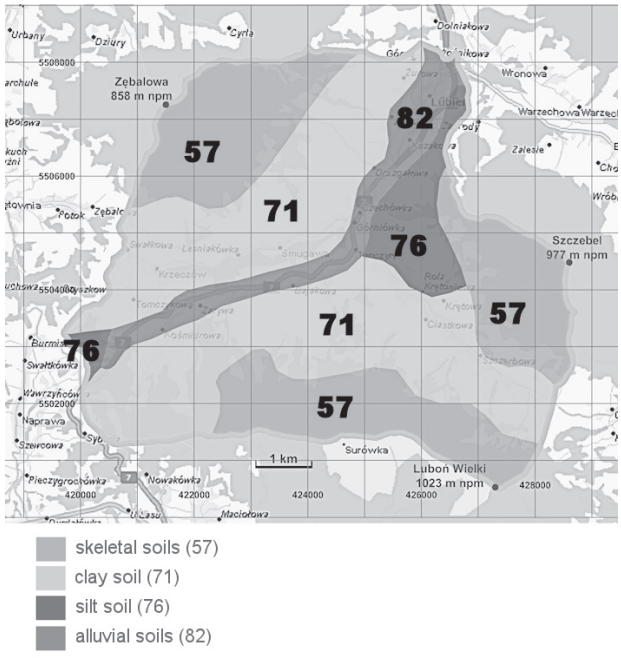


Fig. 5. Soil map
Rys. 5. Mapa gleb

6. Meteorological data

Near all physically-based models requires an actual meteorological conditions in the basin (except fixed data like DEM, channel network, land use, soils). For the CASC2D models it is the height and intensity of precipitation [14].

Precipitation is measured in three meteorological stations located on the watershed area. Each station is equipped with a tipping bucket rain gauge to measure the amount and intensity of precipitation [1]. Stations are located (Fig. 1) in the Krzeczów (station ST01), Lubień (ST02) and Tenczyn (ST03). This station network is a part of designed short-term flood prediction system. In addition there is the fourth station (Luboń Wielki mountain, ST04) belongs to the Polish Institute of Meteorology and Water Management.

Simulations in the CASC2D model require the amount and intensity of precipitation as a spatial map. Since the measurement is performed in a spot, it is necessary to interpolate the data. A Shepard method of interpolation was used [13] (also known as the method of Inverse Distance Weighting) Sample output of the method was presented in Fig. 6.

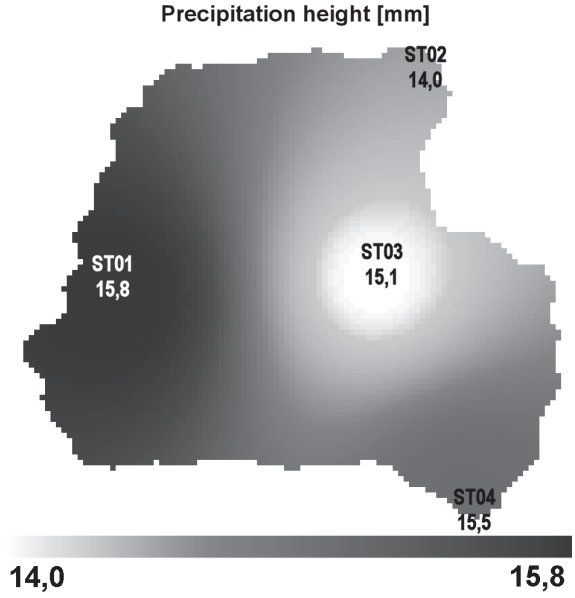


Fig. 6. Precipitation height interpolation using Inverse Distance Weighting method
Rys. 6. Rozkład wysokości opadu metodą Sheparda

7. Flood Prediction System

To design a fully working flood prediction system a mechanisms for the exchange of data between the different elements of the system should be created. The

system is consist of: subsystem of measurement and data transmission, data processing subsystem based on hydrological model and data sharing and presentation subsystem. The entire set of fixed data can be defined a priori in the model, whereas real time data (precipitation) must be delivered to the model without delay. The result of the hydrological model should also be delivered to the end user of the system (i.e. people living on the watershed area).

7.1. System structure and operation

The first subsystem is responsible for data collection. It is a network of automatic measurement stations (maintenance-free). Each station is equipped with a rain gauge, thermometer, hygrometer and a device for data recording and transmission via the GSM mobile network. Each station is powered by a solar panel. Data from stations are transmitted in real time to the database (fig. 7). The database on the server transmits measured data to the data processing subsystem. This is the hydrological model CASC2D. Here, simulations are performed continuously for current weather conditions in the watershed. Simulations determine the predicted outflow from the watershed in the next few hours (the time horizon for the selected watershed was 6 hours).

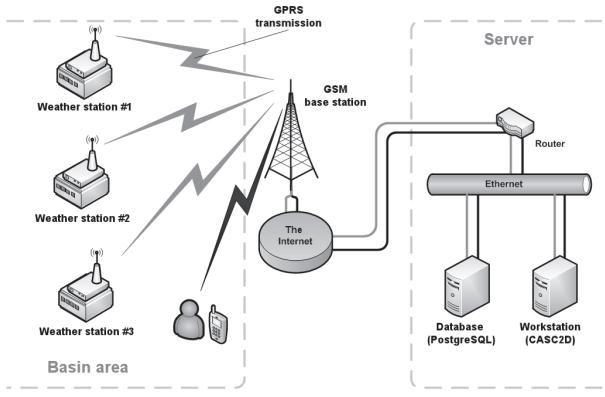


Fig. 7. Data flow in designed system
Rys. 7. Schemat transmisji danych w systemie

The major output of the model is the river discharge and its height in the cross-section of watershed outlet (near Lubień). These data are shown in the data presentation subsystem. Moreover in this subsystem, the measurement data are presented from the individual station.

When the data processing subsystem detects an significant increase of discharge in the forecast period system informs the end user (i.e. people living in affected area). For the communication with users a GSM text message was used.

7.2. Model

The main part of the data processing subsystem is a hydrological model. The basic implementation of CASC2D model is only suitable for simulating a single flood event [4]. As shown simulations performed by author [11], the usefulness of the CASC2D model for continuous simulation is low. It is caused by neglecting long-term effects like subsurface flow and/or evapotranspiration. Without these phenomena it is not possible to create automatic hydrological forecasts .

The basic version of the CASC2D model takes into account such phenomena as: interception (i.e. water retention on the leaves of the trees), infiltration (percolation of water into the ground) and surface runoff. These phenomena significantly affect the simulations of flash floods. With the continuous simulation significant factor in the water balance – especially during the summer – are evaporation and transpiration (in Poland the amount of evaporated water is up to 85% of rainfall). Also the subsurface flow is significant during long-term simulation.

The flow rate of subsurface flow is much lower than the flow rate of surface runoff and therefore does not need to be taken into account in the case of a single event simulations (like flash floods). In the case of continuous simulation this kind of simplification is not enough [11].

The modification of the system was based on the consideration the phenomenon of evapotranspiration and soil drainage combined with subsurface runoff based on a simplified method of finite elements. It was based on the physical interpretation of the phenomena occurring in the watershed. It was possible to automate the model by creating a procedure for determining initial conditions of the model (i.e. initial soil moisture).

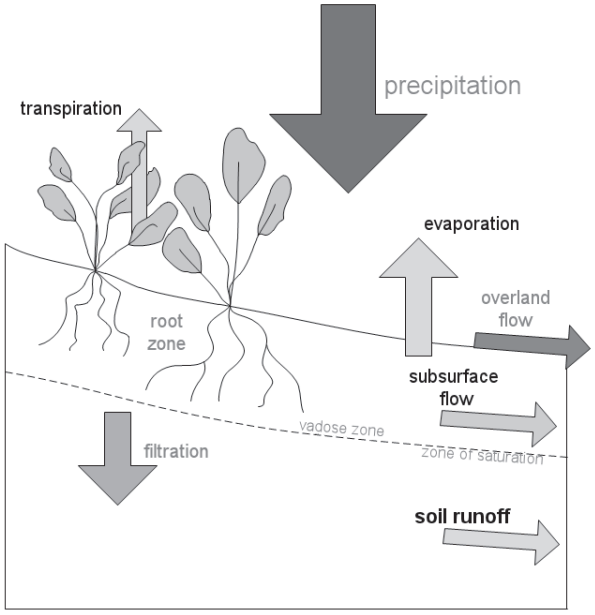


Fig. 8. Phenomena in modified system using CASC2D model

Rys. 8. Zjawiska fizyczne uwzględnione w systemie wykorzystującym modelu CASC2D

7.3. Hydrological forecast

During the flood, which occurred at night of 29th to 30th June 2011 a test of the system in terms of short-term forecast was performed. The maximum intensity of precipitation exceeded 70 mm/h. The rain lasted only 20 minutes and the total precipitation height for each station was approximately 20 mm for the period indicated in Fig. 9. This precipitation event was the direct cause of the flash flood on Lubieńka river.

For testing purposes, it was assumed that the end user will receive information about all the projected flood forecasts on river with gauge height higher than 94 cm (outlet cross-section). The first result of the simulation takes into account a rainfall that occurs at 1:30AM (Table 1). According to the system output the culmination of the flood was expected at 3:30AM. Immediately text message was sent to users with information about the expected flood. The time advance of prediction was approximately two hours. The system as a result

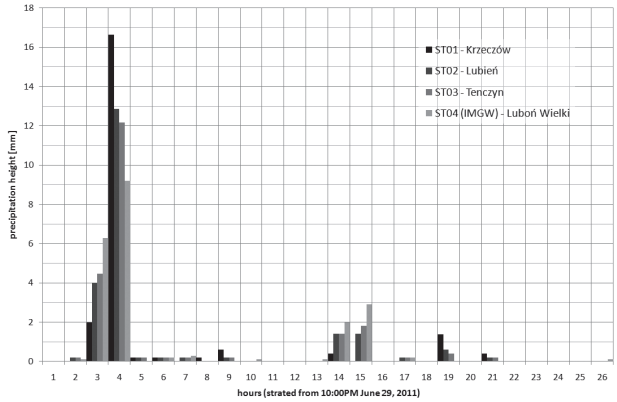


Fig. 9. Precipitation height derived from rain gages

Rys. 9. Wysokość opadu zarejestrowana w stacjach pomiarowych

Table 1. Summary of peak flow forecasting on June 30, 2011 (with received SMS)

Tab. 1. Zastawienie informacji dotyczących przewidywanego wezbrania (kolejne otrzymywane SMSy ostrzegawcze) – dla wezbrania z dnia 30 czerwca 2011

Time of model execution	Time of receiving a text message	Time of predicted culmination wave (as shown in text message)	Predicted outflow [m^3/s]	Predicted river height [cm]	Culmination wabe time advance (3:30 AM) [hour:minute]
01:30	01:34	03:30	8,68	**114**	01:56
01:45	01:49	03:30	8,74	**114**	01:41
02:00	02:04	03:30	8,82	**114**	01:26
02:15	02:19	03:30	8,91	**115**	01:11
02:30	02:34	03:30	8,95	**115**	00:56
02:45	02:49	03:30	8,95	**115**	00:41
03:00	03:04	03:30	8,95	**115**	00:26
03:15	03:19	03:30	8,95	**115**	00:11

of the projected river height at the outlet shows around 115 cm, which corresponded to the actual measurements. The time of culmination wave has been predicted. Fig. 10 shows the outflow hydrograph based on simulation performed in the system with CASC2D model during his operation. Individual text message (vertical lines), which were received by the users was marked also on Fig. 10. The first text message was: "For 1 hour 56 min. (2011.06.30 3:30) Lubieńka river height reaches 114 cm." Subsequent messages differs by information about the time of culmination wave.

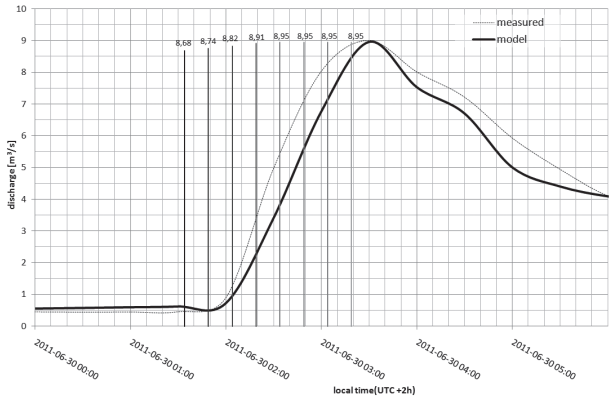


Fig. 10. Hydrological forecast results and flow measurement

Rys. 10. Porównanie prognozy otrzymanej z systemu z przepływem zarejestrowanym

As shown it is possible to predict the flood of nearly two hours in advance. This situation is possible only during torrential rain. Fig. 10 presented runoff hydrograph during the described flood. This is a comparison of the system simulations and measurement of the river height with a gage installed in the watershed outlet cross-section (connected to the station ST02).

8. Conclusion

This paper describes a short-term flood prediction system and describes his principles.

The flood forecast is created with a maximum time advance that is determined by actual weather data. For the test watershed (near 50 km^2) is it about two hours. For larger watersheds with a correspondingly larger measurement network the forecast time increases.

Presented system needs a relevant data divided into two categories: set of fixed data and real time data. Meteorological data (real time) must be supplied to the system by the measuring devices. Fixed data must be developed and validated before starting the system. In this publication author focused on the potential use of geographic data available free of charge.

Summarize the short term flood prediction system works as expected. With a relevant data it can be applied to small watershed and predict a flash floods.

References

[1] Chmielowski W., Nering K. 2011. *System telemetryczny wybranych wielkości hydrometeorologicznych*, Czasopismo Techniczne Politechniki Krakowskiej 2-Ś.

[2] Johnson B.E., Julien P.Y. 2000. *The two-dimensional upland erosion model CASC2D-SED*, The Hydrology-Geomorphology Interface: Rainfall, Floods, Sedimentation, Land use, pp. 107–125.

[3] Jorgeson J., Julien P. 2005. *Peak Flow Forecasting with Radar Precipitation and the Distributed Model CASC2D*, International Water Resources Association, Water International 30, pp. 40–49.

[4] Julien P.Y., Saghafian B., Ogden F.L. 1995. *Raster-Based Hydrologic Modeling of Spatially-Varied Surface Runoff*, American Water Resources Association, Water Resources Bulletin 31.

[5] Mojski J.E. 1972. *Mapa geologiczna Polski*, Wydawnictwa Geologiczne.

[6] NASA, *CASC2D: Hydrologic Modelling*, 2012 http://gcmd.nasa.gov/records/CASC2D.html, [dostęp: 11 lipca 2012].

[7] NASA Jet Propulsion Laboratory *ASTER Global Digital Elevation Map Announcement*, 2012.

http://asterweb.jpl.nasa.gov/gdem.asp [dostęp: 11 lipca 2012].

[8] NASA Jet Propulsion Laboratory *SRTM – Shuttle Radar Topography Mission*, 2009. http://www2.jpl.nasa.gov/srtm/ [dostęp: 11 lipca 2012].

[9] Nering K. 2009. *SRTM – technologia obrazowania powierzchni Ziemi*, Czasopismo Techniczne Politechniki Krakowskiej 1-Ś.

[10] Nering K. 2011. *System GRASS – możliwości i zastosowania.* Część II, Czasopismo Techniczne Politechniki Krakowskiej 2-Ś.

[11] Nering K. 2011. *Mobilny system informacji hydrologicznej on-line z wykorzystaniem zmodyfikowanego modelu CASC2D i architektury klient-serwer*. Praca doktorska, Politechnika Krakowska, Kraków.

[12] Neteler M., Mitasova H. 2008. *Open Source GIS. A GRASS GIS Approach*, Springer, New York.

[13] Shepard D. 1968. *A two-dimensional interpolation function for irregularly-spaced data*, ACM National Conference 1968, pp. 517–524.

[14] Singh V.P., Frevert D.K. 2002. *Mathematical models of small watershed hydrology and applications*, Water Resources Publication, USA.

[15] Singh V.P., Frevert D.K. 2006. *Watershed Models*, Taylor & Francis Group.

GEOINFORMATICA POLONICA
13: 2014

ROBERT OLENIACZ, MATEUSZ RZESZUTEK[1]
oleniacz@agh.edu.pl, rzeszut@agh.edu.pl

ASSESSMENT OF THE IMPACT OF SPATIAL DATA ON THE RESULTS OF AIR POLLUTION DISPERSION MODELING

Key words:

geographic data, air pollutant modeling, CALPUFF, CALMET

Abstract

Advanced dispersion models, taking into account information on the relief and land cover, as well as temporal and spatial variability of meteorological conditions, are beginning to play an increasingly important role in the assessment of the impact on the air quality. There are numerous spatial databases which can be used in this type of a calculation process, however, there is no answer to the question of how the use of appropriate data set of terrain characteristics affects the results of the distribution of air pollutant concentrations at the surface of the ground. This paper presents two different sets of spatial data of the relief and land cover. Then, their impact on the results of modeling the propagation of pollutants in the ambient air was characterized, using the meteorological processor CALMET and the dispersion model CALPUFF. The obtained results of concentrations in the adopted calculation area were compared on the basis of statistical indicators used to assess pollution dispersion models contained in the statistical package BOOT Statistical Model Evaluation Software Package Version 2.0. The obtained results of calculations of the maximum 1-hour concentrations, the maximum 24-hour mean concentrations and annual mean concentrations for the prepared computational grids with a resolution of 1 × 1 km were analyzed.

OCENA WPŁYWU DANYCH PRZESTRZENNYCH NA WYNIKI MODELOWANIA ROZPRZESTRZENIANIA SIĘ ZANIECZYSZCZEŃ W POWIETRZU

Słowa kluczowe:

dane geograficzne, modelowanie zanieczyszczenia powietrza, CALPUFF, CALMET

Abstrakt

W ocenie wpływu na jakość powietrza coraz większą rolę zaczynają odgrywać zaawansowane modele dyspersji, uwzględniające informację o ukształtowaniu i pokryciu terenu, a także zmienność czasową i przestrzenną warunków me-

[1] AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering, Department of Environmental Management and Protection.

teorologicznych. Istnieje wiele przestrzennych baz danych, które można zastosować w tego typu procesie obliczeniowym, jednak brak jest odpowiedzi na pytanie, jak użycie odpowiedniego zestawu danych cech terenu wpływa na wyniki rozkładu stężeń zanieczyszczeń w powietrzu przy powierzchni terenu. W niniejszym opracowaniu przedstawiono dwa odmienne zestawy danych przestrzennych rzeźby i pokrycia terenu. W następnej kolejności scharakteryzowano ich wpływ na wyniki modelowania propagacji zanieczyszczeń w powietrzu atmosferycznym, stosując procesor meteorologiczny CALMET oraz model dyspersji CALPUFF. Porównania uzyskanych wyników stężeń w przyjętym obszarze obliczeniowym dokonano opierając się na wskaźnikach statystycznych stosowanych do oceny modeli rozprzestrzeniania się zanieczyszczeń zawartych w pakiecie statystycznym BOOT Statistical Model Evaluation Software Package Version 2.0. Analizie poddano otrzymane wyniki obliczeń stężeń maksymalnych jednogodzinnych, maksymalnych średniodobowych i średniorocznych dla przygotowanych siatek obliczeniowych o rozdzielczości 1×1 km.

1. Introduction

In the recent years, access to spatial databases has improved significantly. A great portion of these databases is available free of charge for non-commercial, but also commercial, uses. A broader characteristics of available geographic data, as well as the possibilities of their use, can be found in [1]. Spatial databases are also applicable in the system of environment quality management, particularly in the field of assessing impacts of emission sources on the air quality, using techniques of mathematical modeling of pollution dispersion in the air. In this type of calculations, carried out using the latest generation of models, numerical terrain models Global 30 Arc-Second Elevation (GTOPO30) [2], the Shuttle Radar Topography Mission (SRTM) [3] can be used, as well as the land cover data Global Land Cover Characterization (GLCC) [4], made available by the United States Geological Survey (USGS). The above mentioned data are of global character, but not all of them cover the entire globe (e.g. SRTM). There is a possibility to implement other spatial data sets through a system of input data preparation preprocessors of the meteorological model CALMET, but most often they merely cover the area of the United States of America. The methodology of preparing data on the relief and land cover for the meteorological model based on geophysical preprocessors of the CALMET/CALPUFF modeling system, has been described in detail in [5–8].

The capabilities of geophysical preprocessors of the analyzed modeling system are limited and do not fully use the available resources of spatial databases. Using geographic information management tools, such as Quantum GIS or ArcGIS, it is possible to prepare computational grids of terrain characteristics for the CALMET/CALPUFF model. For example, the implementation of Corine Land Cover 2006 (CLC 2006) data for the countries of the European Union can be performed using the above-mentioned geoinformation tools, on the basis of the methodology described in [9]. CLC 2006 land cover data are available free of charge via the European Environment Agency [10] and the Chief Inspectorate of Environmental Protection [11].

In the CALMET and CALPUFF models, the impact of terrain characteristics is considered in two scales: the macro scale (meteorological processor) and the micro scale (processor of dispersion of air pollutants) [7, 8]. The meteorological processor CALMET, in the calculations of the three-dimensional grid of meteorological parameters, takes into account the kinematic effect of a terrain on the wind field, the blocking effect of obstacles, as well as the flows and inflows of air masses from the hills and up the slopes [8]. In the non-steady-state, multi-layer Gaussian model of the CALPUFF puff, four methods can be used of taking into account the impact of the terrain relief on the results of the air pollution dispersion modeling, associated with the preparation of spatial data. These include: NA – No Adjustment method, ISC – Industrial Source Complex Model, PPCT – Plume Path Coefficient Treatment and SCTA – Simple CALPUFF Terrain Adjustments [6]. The set of CALMET/CALPUFF models finds its application in air pollution forecasting systems FAPPS (Forecasting of Air Pollution Propagation System) [12–14] and AIRPOMERANIA

(Regional information management system of air quality in Pomerania) [15, 16], as well as in Environment Protection Programs [17, 18], or in assessing the impact of industrial facilities on the air quality [19–24].

The purpose of this article is to examine the impact of various spatial data on the results of dispersion modeling of pollutants in the ambient air. The analysis was performed for the obtained results of spatial distribution of the maximum 1-hour concentrations, the maximum 24-hour mean concentrations and annual mean concentrations in 676 calculation receptors for two different sets of spatial data of the relief and land cover with a resolution of 1×1 km. Comparison of the models was performed based on the statistical indicators used to evaluate the models of pollutant dispersion. These indicators are collected in the statistical package BOOT Statistical Model Evaluation Software Package Version 2.0, constituting a basic part of the KIT validation model, which was created as a result of a series of conducted workshops and a conference on "Harmonization within Atmospheric Dispersion Modeling for Regulatory purposes" [25–28]. Analysis of the results was limited to six statistical indicators recommended for use in the evaluation of environmental models by the U.S. Environmental Protection Agency (U.S. EPA) [29].

1. Methodology

1.1. Methodology for the preparation of spatial data

In the process of pollution dispersion modeling with the Gaussian puff model, the first step is to create computational grids of geophysical parameters from the available data on the relief and land cover. The preprocessors of the data preparation on digital terrain model – TERREL and land cover – CTGPROC were used. The MAKEGEO preprocessor, based on the previously prepared data, forms a computational grid, containing additional information on the geophysical factors dependent on land cover classes. For the purposes of this analysis, two rectangular computational grids of the sides measuring 26×26 km and a resolution of 1 km were prepared in the coordinate system WGS 84 (World Geodetic System '84), with the UTM map projection (Universal Transverse Mercator), zone 34. The coordinates of the lower left corner x = 420.000, y = 5532.000 were adopted. The grids were designated with respect to the data source with the following symbols: GG (GTOPO30 and GLCC data), SC (SRTM3 and CLC 2006 data). The created computational grids of the terrain relief and land cover classes have been presented in Figures 1–3.

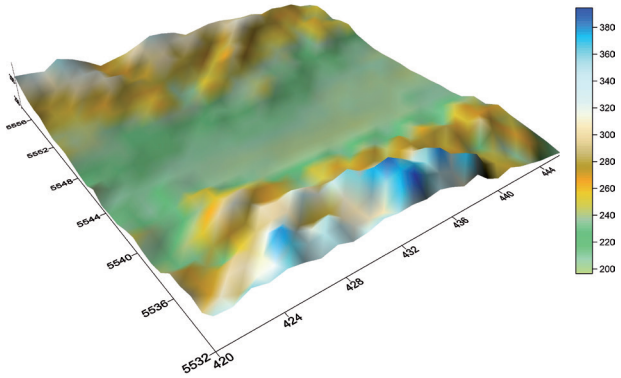


Fig. 1. Three-dimensional visualization of the GTOPO30 digital elevation model for a computational grid with a resolution of 1.0 km

Rys. 1. Trójwymiarowa wizualizacja numerycznego modelu terenu GTOPO30 dla siatki obliczeniowej o rozdzielczości 1,0 km

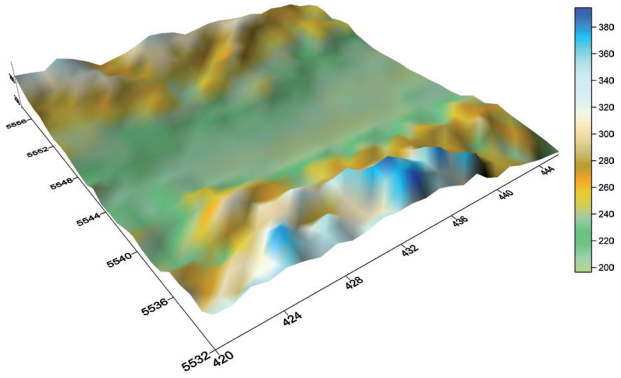


Fig. 2. Three-dimensional visualization of the SRTM digital elevation model for a computational grid with a resolution of 1.0 km

Rys. 2. Trójwymiarowa wizualizacja numerycznego modelu terenu SRTM dla siatki obliczeniowej o rozdzielczości 1,0 km

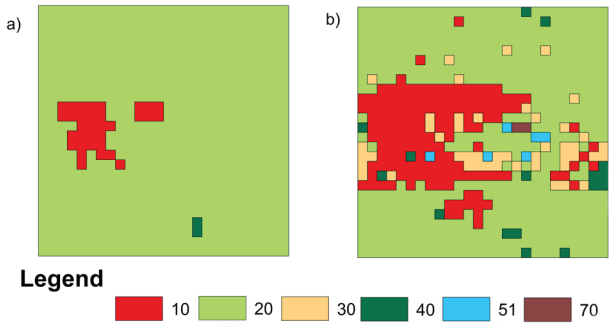


Fig. 3. Graphical representation of the adopted computational grids of the land cover classes: a) GG, b) SC (10 – urban of built-up land, 20 – agricultural land, 30 – rangeland, 40 – forest land, 51 – water, 70 – barren land)

Rys. 3. Graficzne przedstawienie przyjętych siatek obliczeniowych klas pokrycia terenu: a) GG, b) SC (10 – tereny miejskie, 20 – tereny rolnicze, 30 – łąki/pastwiska, 40 – lasy, 51 – woda, 70 – nieużytki)

1.2. Methodology for the preparation of meteorological data

An essential stage of air pollution dispersion modeling is the creation of a number of data input files containing temporal variability of meteorological factors and the performance of the calculations of meteorological conditions variable in time and space, in a three-dimensional grid. The CALMET model has high demands with respect to the input data. It requires the preparation of data from ground-based meteorological station (wind speed and direction, cloud base height, cloudiness, temperature, humidity, pressure and precipitation code) and those from radar surveys, providing information on the variability of the wind field, temperature, pressure and of geopotential height in the vertical section of the atmosphere.

In order to prepare the information from ground-based meteorological stations for the CALMET model, the SMERGE preprocessor was used. This program creates input database files of the meteorological factors with a defined temporal resolution. The data from five stations were used of the period from 1 January 2012 to 31 December 2012. The characteristics of the ground-based meteorological stations from which the meteorological data were used to create the three-dimensional computational grid, have been presented in Table 1.

The data from the radar surveys were processed with the meteorological preprocessor of the vertical structure of the atmosphere READ62. This program for each upper air station creates a separate input data file of the CALMET model. This paper makes use of the information from four upper air stations. The data from these stations have different temporal resolutions; depending on the station and the adopted calculation period, intervals between consecutive measurements can be 6, 8 or 12 hours. The characteristics of the upper air stations used in the calculation of the three-dimensional meteorological grid have been summarized in Table 2.

Table 1. Characteristics of surface meteorological stations

Tab. 1. Charakterystyka powierzchniowych stacji meteorologicznych

USAF[2]	Location (city)	X [km]	Y [km]	Station elevation [m]	Height of the anemometer [m]
12560	Katowice	359.496	5566.061	284	14
12566	Kraków Balice	414.143	5548.215	237	14
12600	Bielsko-Biała	358.078	5519.758	399	14
12575	Tarnów	498.783	5541.668	209	14
12660	Nowy Sącz	477.563	5497.330	295	14

[2] Air Force Datsav3 station number.

Table 2. Characteristics of upper air stations
Tab. 2. Charakterystyka aerologicznych stacji meteorologicznych

Number of station	Location (city)	X [km]	Y [km]	Station elevation [m]
11952	Poprad Ganowce	449,562	5431,020	706
12425	Wrocław	215,833	5744,603	122
11520	Praga Libus	31,000	5560,900	303
12374	Legionowo	497,297	5805,530	96

The previously created geophysical and meteorological data sets are input information of the CALMET model. The meteorological processor, on the basis of the prepared data, forms a three-dimensional grid of the wind field and temperature, and a two-dimensional grid of atmospheric stability classes, Monin-Obukhov length, the height of the mixing layer, friction speed, convective velocity and rate of atmospheric precipitation. Additionally, the following parameters are assigned at the station: temperature, air density, short-wave radiation, relative humidity and precipitation code. As part of this work, two meteorological grids were prepared for the year 2012, differing only in source of the spatial data (GG and SC).

1.3. Variants of calculation regarding transport of pollutants in the air

The calculations of pollution dispersion of in the ambient air were carried out using the multi-layer CALPUFF model. To simplify, these calculations included the emission of one substance only (nitrogen dioxide), assumed at a constant level, and occurring from two twin point emitters having the following dimensions: height 80 m, diameter 2.6 m, located in the center of the adopted computational grid. The calculations were carried out for four variants, two for each geophysical data grid of the terrain. In order to analyze the impact of the applied spatial data of the terrain relief and land cover on the modeling results, for each of the previously prepared computational grid, calculations were performed using the spatial distributions of concentrations with algorithms without taking into account the effects of the relief (NA – no adjustment) and the equation of the simplified CALPUFF method (SCTA – Simple CALPUFF terrain adjustment). The adopted variants of calculation were characterized in Table 3.

1.4. Characteristics of statistical indicators

Comparative analysis was performed, based on numerical relationships in the form of six statistical indica-

Table 3. Summary of variants of calculation of air pollution dispersion
Tab. 3. Zestawienie wariantów obliczeń dyspersji zanieczyszczeń powietrza

Designation	Database terrain elevation	Database land Cover	Terrain adjustment method
GG_NA	GTOPO30	GLCC	No Adjustments (NA)
GG_SCTA	GTOPO30	GLCC	Simple CALPUFF Terrain Adjustments (SCTA)
SC_NA	SRTM3	CLC 2006	No Adjustments (NA)
SC_SCTA	SRTM3	CLC 2006	Simple CALPUFF Terrain Adjustments (SCTA)

tors recommended by the U.S. Environmental Protection Agency (U.S. EPA) to evaluate the models of pollutant propagation in the ambient air. Individual indicators have been presented with the formulas 1–6 and their broader description can be found in [29].

- Fractional Bias (FB):

$$FB = \frac{\left(\overline{C_O} - \overline{C_P}\right)}{0.5 \cdot \left(\overline{C_O} + \overline{C_P}\right)} \quad (1)$$

- Geometric Mean bias (MG):

$$MG = \exp\left(\overline{\ln C_O} - \overline{\ln C_P}\right) \quad (2)$$

- Normalized Mean Square Error (*NMSE*):

$$NMSE = \frac{\overline{\left(C_O - C_P\right)^2}}{\overline{C_O} \cdot \overline{C_P}} \quad (3)$$

- Geometric Variance (VG):

$$VG = \exp\left[\overline{\left(\ln C_O - \ln C_P\right)^2}\right] \quad (4)$$

- Correlation Variance (R):

$$R = \frac{\left(C_O - \overline{C_O}\right)\left(C_P - \overline{C_P}\right)}{\sigma_{C_P} \cdot \sigma_{C_O}} \quad (5)$$

- FAC2:

$$FAC2 = \text{fraction of data that satisfy } 0.5 \le \frac{C_P}{C_O} \le 2 \quad (6)$$

where:
C_P – denotes model predictions,
C_O – denotes observations,
¯(upper line) – represents the arithmetic mean,
σ_c – denotes the standard deviation over the dataset.

The indicators defined above have been used to assess individual results of modeling the levels of pollutant concentrations at the surface of the ground, for selected variants of calculation with regard to the maximum 1-hour concentrations (C_m), the maximum 24-hour concentrations (C_d) and the annual mean concentrations (C_a) of nitrogen dioxide. The relationships between the calculated results obtained without taking into account the terrain relief (NA), and using a simplified CALPUFF method (SCTA) in both computational grids (GG and SC) were analyzed, which were designated as the V.1 case (a comparison of the GG_NA variant with the GG_SCTA variant) and the V.2 case (a comparison of the SC_NA variant with the SC_SCTA variant). Particular attention was focused on examining the differences that exist between the computational grids GG and SC using the method of SCTA designated as the V.3 case (a comparison of the GG_SCTA variant with the SC_SCTA variant). Individual analyzed cases have been summarized in Table 4.

Table 4. Summary of the analyzed cases of a comparative analysis of calculation results
Tab. 4. Zestawienie rozpatrywanych przypadków analizy porównawczej wyników obliczeń

Cases	Variant of the reference	Variant of the analyzed
V.1	GG_NA	GG_SCTA
V.2	SC_NA	SC_SCTA
V.3	GG_SCTA	SC_SCTA

2. Analysis of the results

Table 5 summarizes the mean, maximum and minimum values, as well as the standard deviation of the obtained results of calculations of the maximum 1-hour concentrations (C_m), the maximum 24-hour concentrations (C_d) and the annual mean concentrations (C_a) in the adopted area of calculations, divided into individual analyzed variants of calculation.

The presented values show that in the case of performing calculations using the simplified CALPUFF method (SCTA), there is an increase in the value of the maximum and mean concentrations in the air. Moreover, the obtained results are characterized by a greater scatter in the calculation area in relation to a variant not taking into account the terrain relief (NA), in all the considered averaging times and the applied computational grids. Higher concentration values of the analyzed substance (NO_2) in the atmospheric air, depending on the adopted computational grid and the time averaging the results of calculations ranged from 3% to 34%. Minimum concen-

Table 5. Summary of minimum, maximum and mean values, as well as of standard deviation for the calculation results of the concentrations C_m, C_d and C_a, obtained in the adopted computational grids and variants of calculation [µg/m^3]

Tab. 5. Zestawienie wartości minimalnych, maksymalnych, średnich i odchylenia standardowego dla wyników obliczeń stężeń C_m, C_d i C_a uzyskanych w przyjętych siatkach i wariantach obliczeniowych [µg/m^3]

Concentration	Grid computational	Minimum	Maximum	Average	Standard deviation
C_m	GG_NA	0.96774	55.7370	8.4191	6.6682
	GG_SCTA	1.12830	71.7070	11.2873	7.3347
	SC_NA	0.88105	52.1520	9.2632	6.3986
	SC_SCTA	0.80429	60.3790	10.4080	7.0158
C_d	GG_NA	0.09362	4.1921	0.9874	0.6298
	GG_SCTA	0.09743	4.4163	1.2082	0.6769
	SC_NA	0.07199	5.9692	1.1057	0.6901
	SC_SCTA	0.07005	7.8337	1.1854	0.7631
C_a	GG_NA	0.00576	0.5142	0.0885	0.0770
	GG_SCTA	0.00607	0.5316	0.0999	0.0812
	SC_NA	0.00523	0.8502	0.1065	0.0953
	SC_SCTA	0.00518	0.8956	0.1109	0.1000

trations obtained in the considered computational receptors of the grid GG (GTOPO30 and GLCC) are characterized by an identical relationship. In contrast, the minimum concentration values determined for the grid SC (SRTM3 and CLC 2006) show the opposite trend. Concentration minima obtained for the SCTA method are even lower by 9% than the minimum values obtained in the NA method. The use of algorithms defining the impact of terrain relief on the dispersion of air pollutants, in the calculations is revealed as an increase in the maximum and mean concentrations obtained in the analyzed computational grids. Analysis of the results carried out for the assumed variants of calculation: GG_SCTA and SC_SCTA, shows that the computational grid SC has lower concentration values, which additionally are characterized by a lower scatter around the mean value. This situation was particularly evident for the maximum 1-hour concentrations obtained in the calculations. It should also be noted that the highest maximum daily mean concentrations and annual mean concentrations occurred in the computational grid SC, and their difference with respect to the grid GG amounted to 77 and 68%, respectively.

Table 6 summarizes the values of statistical indicators of environmental model evaluation respective of the obtained results of the concentrations C_m, C_d and C_a in the adopted calculation area, and with a division into the analyzed V.1, V.2 and V.3 cases, characterized in Table 4.

The systematic error values (FB) for the V.1 and V.2 cases, presented in Table 6, unambiguously prove that the application of the method not taking into account the terrain relief (NA) in the process of pollution dispersion modeling is characterized by underestimation of the results of concentrations in the air in all the considered averaging times. This finding is confirmed by linear graphs of systematic error variation (FB) in the analyzed 676 point receptors, contained in drawings 4–6. Additionally, in each considered averaging time for the V.1 and V.2 cases, at least 74% of the results of concentrations were underestimated in relation to the sim-

Table 6. Summary of statistical indicators of assessing environmental models for the considered V.1, V.2 and V.3 cases, relative to the obtained results of the concentrations C_m, C_d and C_a in the adopted calculation area

Tab. 6. Zestawienie wartości wskaźników statystycznych oceny modeli środowiskowych dla rozpatrywanych przypadków V.1, V.2 i V.3 ze względu na uzyskane wyniki stężeń C_m, C_d i C_a w przyjętym obszarze obliczeniowym

Indicator	C_m			C_d			C_a		
	V.1	V.2	V.3	V.1	V.2	V.3	V.1	V.2	V.3
FB	-0.291	-0.116	0.081	-0.201	-0.070	0.019	-0.121	-0.040	-0.105
MG	0.700	0.890	1.100	0.790	0.940	1.080	0.850	0.960	0.960
NMSE	0.230	0.050	0.170	0.100	0.020	0.120	0.030	0.010	0.100
VG	1.340	1.050	1.170	1.120	1.010	1.060	1.040	1.000	1.040
R	0.794	0.961	0.801	0.928	0.988	0.913	0.991	0.999	0.984
FAC2	0.814	0.988	0.910	0.953	1.000	0.975	1.000	1.000	1.000

plified CALPUFF method. Different observations were made with the results obtained for the V.3 case, where, depending on the adopted averaging time, the values of mean systematic errors contained in Table 6 indicate an average underestimation or overestimation of the calculation results. The values of the correlation coefficient (R) and of the model quality ratio (FAC2) illustrate that with the increasing averaging time of concentrations, the analyzed cases are characterized by a better correlation and by a smaller number of two-fold overestimation and underestimation of the calculation results with respect to variations of reference. In the case of mean annual concentrations, there was no two-fold overestimation and underestimation of the results in the adopted calculation area (FAC2 = 1.000). The above statements are reflected in the presented values of geometric mean error (MG), geometric mean of variance (VG) and the relative standard deviation (NMSE).

Variability of the systematic error (FB) in the grid receptors, illustrated in Figures 4–6, shows a large variety of the obtained calculation results of pollution dispersion for the V.3 case (a comparison of GG_SCTA with SC_SCTA). At any time of averaging the results of concentrations in the air (C_m, C_d and C_a) for this case, there is both a large number of FB values indicating the underestimation and overestimation. The biggest discrepancies have been presented in Figure 4, showing the variability of FB of the analyzed cases with respect to the maximum 1-hour concentrations. A comparison of the obtained FB values for the V.3 case proves that, the underestimation and overestimation of the calculation results was the case with, respectively, 58.3 and 41.7% of the grid receptors of the concerned calculation area. On the other hand, 7.1% of the results were underestimated two-fold, and 1.9% of the results were overestimated. Based on the obtained FB values for the calculations of 1-hour maximum concentrations, it can be concluded that the results of the variant of calculation SC_SCTA are more frequently underestimated, when compared with the GG_SCTA variant. This finding is reflected in the FB values obtained for the maximum 24-hour concentrations, where the underestimation of the results took place in more than 62% of the receptors of the SC_SCTA variant.

The relationship for the V.3 case presented in Figure 6 proves otherwise. The SC_SCTA variant is distinguished by a larger share of the overestimated results, which occur in the case of 52.4% of the receptors. This is confirmed by the FB mean value (Table 6), which points to a slight underestimation of the results for the GG_SCTA variant in the considered calculation area. Furthermore, the two-fold overestimation and underestimation of the results were not noted in this case.

The values of statistical indicators presented in Table 6, parameterizing the existing relationships between the analyzed variants of calculation, theoretically may prove a good, and even high, convergence of the

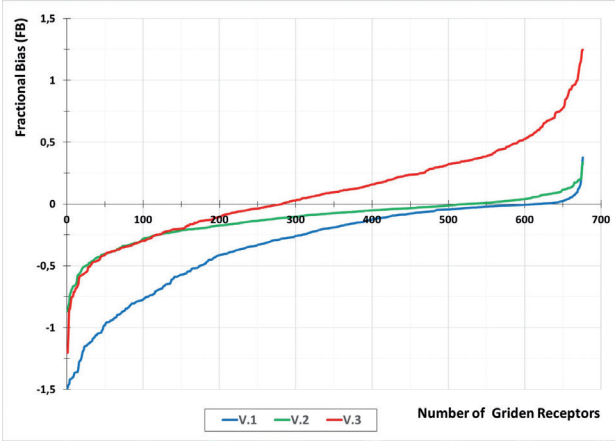


Fig. 4. Three-dimensional visualization of the GTOPO30 digital elevation model for a computational grid with a resolution of 1.0 km

Rys. 4. Trójwymiarowa wizualizacja numerycznego modelu terenu GTOPO30 dla siatki obliczeniowej o rozdzielczości 1,0 km

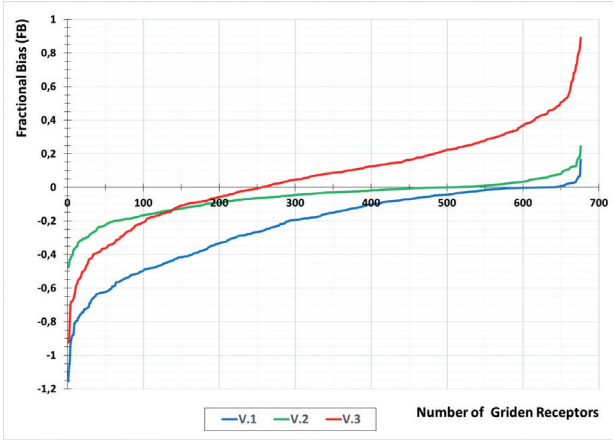


Fig. 5. Graphical representation of the variability of a systematic error (FB) in the computational receptors of the analyzed V.1, V.2 and V.3 cases, relative to the obtained results of calculations of the maximum 24-hour concentrations (C_d)

Rys. 5. Graficzne przedstawienie zmienności wartości błędu systematycznego (FB) w receptorach obliczeniowych analizowanych przypadków V.1, V.2 i V.3 w odniesieniu do uzyskanych wyników obliczeń stężeń maksymalnych 24-godzinnych (C_d)

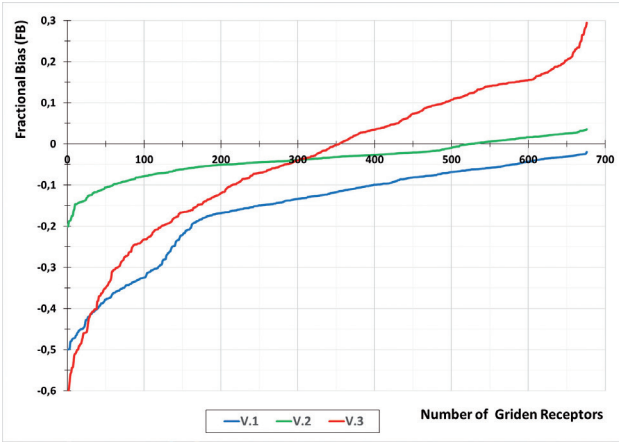


Fig. 6. Graphical representation of the variability of a systematic error (FB) in the computational receptors of the analyzed V.1, V.2 and V.3 cases, relative to the obtained results of calculations of the mean annual concentrations (C_a)

Rys. 6. Graficzne przedstawienie zmienności wartości błędu systematycznego (FB) w receptorach obliczeniowych analizowanych przypadków V.1, V.2 i V.3 w odniesieniu do uzyskanych wyników obliczeń stężeń średniorocznych (C_a)

results for the V.3 case. Additionally, with the increasing averaging times of concentrations in the air, statistical indicators show even stronger correlation between the analyzed variants of calculation. The FB values presented in Figures 4–6 in the various receptors of the adopted calculation area, highlight, however, imprecision of this type of statements. The approximately even distribution of the percentage share of the positive and negative FB values, however, leads to acquisition of inadequate information, indicating a strong correlation between the analyzed variants, although the difference between the obtained results of calculations for the GG_SCTA and SC_SCTA variants are clear, reaching even up to two-fold underestimation and overestimation.

3. Summary

As it is evident from the conducted calculations and analyses, the application of various algorithms which take into account the terrain relief and spatial databases

characterized by different properties, has a significant impact on the results of air pollution concentration modeling at the surface of the ground, which is implemented by means of the CALMET/CALPUFF models.

It was found that including the terrain characteristics into this type of calculation process is reflected by an increase in mean and maximum values in the adopted calculation area, irrespective of the computational grid and averaging time. The values of the model evaluation indicators prove that the application of the SCTA method for the calculation of pollution dispersion is characterized by a strong overestimation of the calculation results of all the considered averaging times (maximum 1-hour and 24-hour concentrations and annual mean concentrations) and of both computational grids (GG and SC).

The statistical indicators which were used in the analysis (FB, MG, VG, NMSE, R and FAC2) provided coherent observations regarding the compared variants of calculation (V.1, V.2 and V.3 cases). With the increasing averaging time of the calculation results of spatial distribution of concentrations of the analyzed gaseous substance (NO_2), individual cases were characterized by a better correlation, as well as by a fewer number of two-fold overestimations and underestimations. The correctness of this statement is perfectly illustrated by the range of occurrence of the FB value, which for the maximum 1-hour and annual mean concentrations is contained within the range: (-1.5; 1.3) and (-0.6; 0.3).

A blind interpretation of the values of statistical indicators that define numerical dependencies between the variants of calculation for the V.3 case may pose some problem. They show a good, and even high, correlation of the calculation results, particularly with respect to the mean annual concentrations. This fact is, however, derived from a relatively even distribution of the overestimated and underestimated results. In this case, more adequate information is provided by the results of systematic error calculations in the various receptors. They clearly reveal a fact that the application of different spatial data is characterized by a large scatter of the values in the case of underestimation and overestimation of the calculation results. This is particularly evident in the case of spatial distribution of the maximum 1-hour concentrations, where in the range of ($-0.1 \leq FB \leq 0.1$), characterized by a high correlation, there are only 22.5% of the calculation results taken into account, and two-fold overestimation and underestimation occurred for up to 9% of the receptors.

The application of spatial data characterized by lower accuracy (GTOPO30, GLCC) in the process of modeling levels of substances in the air, may result in very high discrepancies in the calculation results, especially in the case of maximum 1-hour and mean daily concentrations. However, with the increasing averaging time of calculation results, the values of systematic errors are getting smaller, indicating a better correlation of the variants analyzed in the V.3 case. For the annual mean concentrations, only 9.6% of the obtained results go beyond the boundaries of the so-called good model (FB value = 0.3), already proving a high comparability of these variants.

The paper has been prepared within the scope of the AGH UST statutory research no. 11.11.150.008.

References

[1] Nering K. 2012. Public geospatial data sources. Geoinformatica Polonica, Vol. 12, pp. 63–72.

[2] Global 30 Arc-Second Elevation (GTOPO30), https://lta.cr.usgs.gov/GTOPO30.

[3] Shuttle Radar Topography Mission (SRTM), https://lta.cr.usgs.gov/SRTM2.

[4] Global Land Cover Characterization (GLCC), https://lta.cr.usgs.gov/GLCC.

[5] Earth Tech, Inc. 2006. Development of the Next Generation Air Quality Models for Outer Continental Shelf (OCS) Applications (CALMET and Preprocessors). Concord.

[6] Scire J.S., Strimaitis D.G., Yamartino R.J. 2005. A User's Giude for the CALPUFFF Dispersion Model. Concord.

[7] Scire J.S., Robe F.R., Fernau M.E., Yamartino R.J. 2000. A User's Guide for the CALMET Meteorological Model (Version 5), Concord.

[8] Szczygłowski P. 2007. Ocena przydatności wybranych modeli gaussowskich w obliczeniach stanu zanieczyszczenia powietrza. Dysertacja doktorska, Wydział Geodezji Górniczej i Inżynierii Środowiska AGH, Katedra Kształtowania i Ochrony Środowiska, Kraków (unpublished).

[9] Rzeszutek M. 2013. Ocena oddziaływania Zakładu Termicznego Przekształcania Odpadów Komunalnych w Krakowie na jakość powietrza z wykorzystaniem matematycznego modelu dyspersji CALPUFF. Praca magisterska, Wydział Geodezji Górniczej i Inżynierii Środowiska AGH, Katedra Kształtowania i Ochrony Środowiska, Kraków (unpublished).

[10] European Environment Agency (EEA), http://www.eea.europa.eu/.

[11] CORINE Land Cover (CLC), http://clc.gios.gov.pl.

[12] Hajto M.J., Godłowska J., Kaszowski W., Tomaszewska A.M. 2012. System prognozowania rozprzestrzeniania zanieczyszczeń powietrza FAPPS – założenia, możliwości, rozwój. [In:] Ochrona powietrza w teorii i praktyce, Vol. 2 (ed.) J. Konieczyński. Instytut Podstaw Inżynierii Środowiska PAN, Zabrze, pp. 89–96.

[13] Godłowska J., Kaszowski W., Hajto M.J., Tomaszewska A.M. 2012. Wpływ sposobu przygotowania meteorologicznych danych wejściowych w systemie FAPPS na jakość pola wiatru i głębokości mieszania. Konsekwencje dla prognozy PM10. W: Ochrona powietrza w teorii i praktyce, Vol. 2 (ed.) J. Konieczyński. Instytut Podstaw Inżynierii Środowiska PAN, Zabrze, pp. 75–88.

[14] Forecasting of Air Pollution Propagation System (FAPPS), http://www.smog.imgw.pl /home.

[15] Bielawska M., Kałakowski T., Potrykus R., Stańczyk J. i Szymańska K. 2012. Regionalny system zarządzania informacją o jakości powietrza. [In:] Ochrona powietrza w teorii i praktyce, Vol. 2 (ed.) J. Konieczyński. Instytut Podstaw Inżynierii Środowiska PAN, Zabrze, pp. 23–26.

[16] Regionalny system zarządzania informacją o jakości powietrza w województwie pomorskim – AIRPOMERIANA, http://airpomerania.pl/index.html.

[17] Biuro Studiów i Pomiarów Proekologicznych „EKOMETRIA" Sp. z o.o.: Opracowanie prognozy zanieczyszczenia powietrza pyłem drobnym w Polsce na lata 2010, 2015 i 2020 wraz z analizą uwarunkowań i oceną kosztów osiągnięcia standardów dla pyłu określonych dyrektywą w sprawie jakości powietrza atmosferycznego i czystszego dla Europy. Gdańsk 2009.

[18] ATMOTERM S.A. 2013. Program Ochrony Powietrza dla województwa małopolskiego. Kraków.

[19] Szczygłowski P., Mazur M. 2005. Zastosowanie modelu Calmet/Calpuff do obliczeń poziomu stężeń zanieczyszczeń pochodzących z wysokich emitorów punktowych. Inżynieria Środowiska, Vol. 10, No. 2, pp. 195–205.

[20] Szczygłowski P., Mazur M. 2006. Modelling dispersion of air pollutants over the area of diversified relief based on the Calmet/Calpuff model. Environment Protection Engineering, Vol. 32, No. 4, pp. 73–77.

[21] Levya J., Spenglera J., et al. 2002. Using CALPUFF to evaluate the impacts of power plant emissions in Illinois: model sensitivity and implications. Atmospheric Environment, Vol. 36, pp. 1063–1075.

[22] Tartakovsky D., Broday D.M., Stern E. 2013. Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain. Environmental Pollution, Vol. 179, pp. 138–145.

[23] Ranzatoa L., Baraussea A., Mantovani A., Pittarello A., Benzo M., Palmeri L. 2012. A comparison

of methods for the assessment of odor impacts on air quality: Field inspection (VDI 3940) and the air dispersion model CALPUFF. Atmospheric Environment, Vol. 61, pp. 570–579.

[24] Vieira de Melo A.M., Santos J.M., Mavroidis I., Reis Junior N.C. 2012. Modelling of odour dispersion around a pig farm building complex using AERMOD and CALPUFF. Comparison with wind tunnel results. Building and Environment, Vol. 56, pp. 8–20.

[25] Chang J.C., Hanna R.S. 2004. Air quality model performance evaluation. Meteorol. and Atmos. Phys., Vol. 87, pp. 167–196.

[26] ASTM, Standard guide for statistical evaluation of atmospheric dispersion model performance. American Society for Testing and Materials, Designation D 6589-00. ASTM, 100 Barr Harbor Drive, West Conshohocken, PA 19428–2959.

[27] Olesen H.R. 2005. User's Guide to the Model Validation Kit. National Environmental Research Institute – Ministry of the Environment, Copenhagen

[28] Cox W.M., Tikvart J.A. 1990. A statistical procedure for determining the best performing air quality simulation model. Atmospheric Environment, Vol. 24A, pp. 2387–2395.

[29] Chang J.C., Hanna R.S. 2005. Technical Descriptions and User's Guide for the BOOT Statistical Model Evaluation Software Package, Version 2.0.

GEOINFORMATICA POLONICA
13: 2014

JAN PIEKARCZYK[1]
piekarczyk@agh.edu.pl

APPLICATION OF REMOTE SENSING IN AGRICULTURE

Key words:

remote sensing, satellite images, airborne photographs, spectral properties of vegetation

Abstract

With increasing intensity of agricultural crop production increases the need to obtain information about environmental conditions in which this production takes place. Remote sensing methods, including satellite images, airborne photographs and ground-based spectral measurements can greatly simplify the monitoring of crop development and decision-making to optimize inputs on agricultural production and reduce its harmful effects on the environment. One of the earliest uses of remote sensing in agriculture is crop identification and their acreage estimation. Satellite data acquired for this purpose are necessary to ensure food security and the proper functioning of agricultural markets at national and global scales. Due to strong relationship between plant bio-physical parameters and the amount of electromagnetic radiation reflected (in certain ranges of the spectrum) from plants and then registered by sensors it is possible to predict crop yields. Other applications of remote sensing are intensively developed in the framework of so-called precision agriculture, in small spatial scales including individual fields. Data from ground-based measurements as well as from airborne or satellite images are used to develop yield and soil maps which can be used to determine the doses of irrigation and fertilization and to take decisions on the use of pesticides.

ZASTOSOWANIE TELEDETEKCJI W ROLNICTWIE

Słowa kluczowe:

teledetekcja, obrazy satelitarne, zdjęcia lotnicze, właściwości spektralne roślin

Abstrakt

Przy wzrastającej intensywności roślinnej produkcji rolniczej zwiększa się konieczność pozyskiwania informacji o warunkach środowiskowych, w jakich ta produkcja się odbywa. Metody teledetekcyjne – w tym między innymi obrazy satelitarne, zdjęcia lotnicze i naziemne pomiary spektralne – mogą znacznie ułatwić kontrolę rozwoju roślin uprawnych i podejmowanie decyzji mających na celu maksymalne wykorzystanie nakładów na produkcję rolniczą przy minimalnym, szkodliwym jej wpływie na środowisko. Najwcześniejsze zastosowania teledetekcji w rolnictwie obejmują identyfikację upraw i szacowanie ich areału. Dane satelitarne pozyskiwane w tym celu są niezbędne dla zapewnienia bezpieczeństwa żywnościowego oraz prawidłowego funkcjonowania rynków płodów rolnych w skali krajowej i globalnej. Dzięki ścisłej zależności między ilością promieniowania elektromagnetycznego odbitego od roślin (w określonych zakresach widma)

[1] Adam Mickiewicz University Poznań, Faculty of Geosciences.

i rejestrowanego następnie przez czujniki z pułapu satelitarnego, lotniczego i naziemnego, możliwe jest prognozowanie plonów. Inne zastosowania teledetekcji są bardzo intensywnie rozwijane w ramach tzw. rolnictwa precyzyjnego, w dużych skalach przestrzennych obejmujących pojedyncze pola. Na podstawie pomiarów naziemnych, obrazów satelitarnych i lotniczych tworzone są mapy plonów i mapy właściwości gleb, ustalane są dawki nawodnieniowe i nawożeniowe oraz podejmowane decyzje o stosowaniu chemicznych środków ochrony roślin.

1. Introduction

One of the major global problems is to feed the growing population with increasing pressure on the environment and shrinking arable land resources. Ability to increase food production by increasing the cultivated land area has already been exhausted. The solution would be to increase the intensity of agricultural production in which remote sensing methods can play a crucial role. The introduction of modern methods of agricultural production is likely to require increasing amounts of information concerning conditions under which this production takes place. One way of obtaining this type of environmental data can be applied remote sensing methods, which include ground-based measurements as well as airborne and satellite imaging.

One of the primary and the earliest uses of remote sensing in agriculture is crop identification and area estimation. Such information is necessary to ensure food security at regional and global scale. For yields forecasting remote sensing data are used directly in the regression models, or as input data to agrometeorological models for their calibration. Subsequent applications are very intensively developed in the framework of the so-called precision farming, at small spatial scales including individual fields. Data from ground-based measurements as well as from airborne or satellite images are used to develop yield and soil maps which can be used to determine the doses of irrigation and fertilization and to take decisions on the use of pesticides.

2. Crop identification and area estimation

In many countries, institutions involved in agricultural statistics use remote sensing to identify crops and area estimation. In the USA, the National Agricultural Statistics Service (NASS) uses mainly traditional methods based on information collected from farmers. However, remote sensing methods are introduced increasingly to create land cover maps and acreage estimates with the Cropland Data Layer program (Miller et al. 2009). The most important reason for which the Agency still uses mainly conventional methods of estimating crop area is that satellite images do not meet the time requirements of reports on the estimation of crop acreage, which must be prepared by NASS. The most appropriate time to accurately distinguish corn from soybeans in the upper Midwest occurs around mid-August while the first crop acreage estimates must be published at the end of March in the Prospective Plantings Report and actual plantings are released in the Acreage Report at the end of June. Acreage estimates based on remote sensing data are carried out annually only in seven to ten States on relatively small areas corresponding to the county level. The combination of remote sensing data with ground data results in an error three times smaller in comparison with using only ground-based data.

NASS uses data from the satellites Landsat, MODIS, IRS-P6 and to a lesser extent, Ikonos, Quickbird and AQUA (AMSR). Landsat data are essential for the NASS application but limiting their use is low temporal revisit time of 16 days. A single satellite overpass over the same area within 16 days is not enough for reliable identification of crops in the peak vegetation season because of the high probability of cloud cover. Therefore, it is necessary to combine the use of Landsat 5 and Landsat 7, which will reduce the return visit to 8 days. Each year NASS prepares a set of high quality Landsat images, “best available” for the crops season of interest.

In China, Remote Sensing Application Center (RSAC) of the Ministry of Agriculture monitors in the whole country five main crops: rice, wheat, corn, soybean and cotton. Their area is estimated in the month before a

harvest and in late stages of development, their condition is controlled every month. Monitoring of soil moisture content is conducted once every two weeks. For crop area assessment RSAC mainly adopts the commercial satellite data of Landsat, SPOT, IRS and CBERS. The CBERS program is a partnership between the Chinese and Brazilian governments and include three satellites (currently running two) equipped with three sensors with spatial resolution from 20 m to 260 m and temporal resolution from 3 days to 26 days.

In Europe, the MARS project (Monitoring Agriculture with Remote Sensing) provides timely agricultural production data for the current season at the regional and higher levels. The European Commission needs these information for the implementation of the Common Agricultural Policy and Food Aid and Food Security Policy. The methods of assessing crop area and structure with the use of aerial and satellite images applied in MARS project has been adopted to Polish conditions and agricultural areas were determined on the basis of Landsat Thematic Mapper data (Bochenek et al. 1997). The estimation accuracy of acreage of the main crops in Poland with the use of remote sensing data was very high and this method can be applied by the Central Statistical Office as a supporting and control tool for the traditional methods of agricultural census.

3. Food security

On a global scale remote sensing data are used to control the food situation and predict its change. Production potential of agriculture in the world is very diverse and often does not coincide with the demand for food. Maps of the spatial distribution of carbon uptake by terrestrial plants are developed on the basis of satellite images. The amount of the absorbed carbon by plants is closely related to the size of net primary production and the amount of biomass. The maps compiled with NOAA-AVHRR can be modified by models to indicate what portion of the biomass is consumed by humans. These data together with modeled derivatives are used in combination with ground-based information to better assess potential impacts to the food supply system (Budde et al. 2010).

4. Yield estimations

The use of remote sensing data in agriculture is possible due to variation of the spectral properties of plants and soil. Reflectance from both individual leaves and plant canopies in the visible portion of the spectrum of electromagnetic radiation is low because of the absorption properties of pigments while in the near-infrared (NIR) is high since live vegetation strongly scatter in this wavelengths (Gates *et al*. 1965, Knipling 1970). The transition from low reflectance in the visible region of the spectrum to high NIR reflectance is called "red-edge" and is most useful in agricultural applications (Fig. 1).

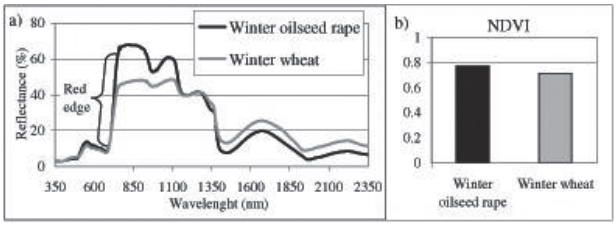


Fig. 1. Spectral signatures a) and NDVI b) for winter oilseed rape and winter wheat recorded during ground measurements on 30 April and 1 May

Rys. 1. Krzywe spektralne a) oraz wartości wskaźnika NDVI b) uprawy rzepaku ozimego i pszenicy ozimej uzyskane na podstawie danych z pomiarów naziemnych wykonanych 30 kwietnia i 1 maja

Spectral curves of winter oilseed rape and winter wheat differ significantly in NIR range due to differences in biomass of plants of these two crops (Piekarczyk 2011). Oilseed rape plants produce more green mass and therefore reflect more NIR radiation. Vegetation indices (VIs) are commonly used to more clearly show the differences in spectral properties of crops. The simple VIs are calculated as ratios or differences of two or more bands in the visible and NIR wavelengths and most widely used is Normalized Difference Vegetation Index (NDVI) (Larsson 2002, Scanlon et al. 2002, Sims and Gamon 2002). The sensors on satellites such as Landsat, SPOT and MODIS, most commonly used in applications for agriculture, record reflectance in a few channels: blue,

green, red, NIR and shortwave infrared (SWIR). Recently launched satellites e.g. WorldView2, RapidEye and Sumbundila Sat have an additional „red-edge" channel over the range from 690 nm to 745 nm.

Total reflectance from crops in the red and NIR wavelengths is very strongly related to plant biomass and the degree of soil cover, which are closely linked with a yield. Knowing the amount of the biomass, determined on the basis of spectral data gathered at different phenological stages, it is possible, with varying degree of probability, to estimate a yield. The relationship between spectral data and yield depends largely on the time in which these data are collected and generally, the strongest correlation occurs when plant biomass is the greatest. In the case of oilseed rape the strongest relationship between NDVI and yield occurs shortly before flowering, in the budding stage, which usually occurs in Wielkopolska region in a third decade of April (Piekarczyk and Sulewska 2011). The green biomass and soil cover do not reach a maximum at this time but in the later stages the correlation between NDVI and yield weakens due to the increase reflectance in visible wavelengths from flowers and saturation of NIR reflectance. A significant problem in estimating crop yields in the later stages of development is the saturation of the reflectance or VIs, when despite increasing biomass their values do not change.

Wheat plants have much lower biomass and cover soil to a lesser extent than oilseed rape plants. The relationship between wheat yield and NDVI is strongest at the stage of shooting (on average in mid-May in Wielkopolska) when plant biomass is at its maximum in a season but less than in case of oilseed rape and therefore the saturation of spectral data does not occur. The dynamics of crops growth and their bio-physical parameters such as biomass, soil cover, leaf area index (LAI) and concentration of chlorophyll are determined by a date of sowing. Winter oilseed rape plants are very sensitive to sowing date and its shift by one week can cause statistically significant differences in yield. Figure 2 shows clear spectral differences between oilseed rape crops sown on different dates (Piekarczyk and Sulewska 2011). In Wielkopolska optimum sowing period for that crop lasts until 25 August (Tys et al. 2003) and the reflectance from the crop sown on this date is higher in NIR wavelength compared with crops sown earlier or later.

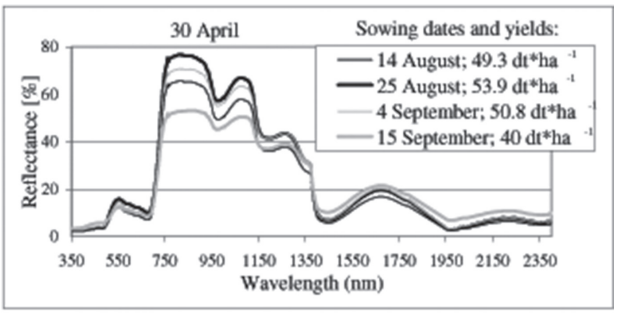


Fig. 2. Spectral signatures for winter oilseed rape sown on four dates

Rys. 2. Krzywe spektralne upraw rzepaku ozimego wysianych w czterech terminach

In Poland a method for estimating yields of the major cereal crops on a national scale was developed. In this method two vegetation indices, Vegetation Condition Index and Temperature Condition Index calculated from low-resolution satellite images for the whole country at weekly intervals in 14 years were used. Values of these indices obtained early in spring (mid April) and early in summer (June) were applied in a yield prediction model. The yield estimations were then compared with data provided by the Central Statistical Office and the average error was 3.83% (Dąbrowska-Zielińska et al. 2002).

5. Precision agriculture

Many types of remote sensing systems are used in agriculture but the most common is a passive system in which the electromagnetic energy reflected from plants is recorded. However much attention is given to the combining traditional spaceborne optical data from the visible and infrared wavelengths with the longer wavelengths of radar in order to reduce the impact of cloud cover. Sensors are mounted on tractors, manned and unmanned aircrafts and on satellites. The recorded data are transmitted to state institutions or private companies involved in agricultural consultancy, where, after

processing and correction they can be used as input data for models in decision support systems.

The management of agricultural crop production requires information on seasonally stable and variable conditions, as well as information for explanation of the causes of yield variability (Moran *et al*. 1997). Stable conditions such as relief, crop yields, soil fertility and its particle size distribution in certain parts of the field are constant during the growing season and do not change in subsequent years. On the basis of these information maps of crop yields and soil properties are compiled. Equally important is information on the variable conditions which change continually during the growing season, such as plant disease, insect and weed infestation and soil moisture. Images obtained by the sensors mounted on aircrafts and satellites can provide knowledge about the entire field and not just the places where samples were taken or traditional ground observations were made. Using information on both the constant and variable conditions in a season it is possible to obtain information about the causes of yield variation and the possibility of undertaking activities that ensure an optimum yield at given conditions. For example, information about the state of the drainage network can be based on information about the variation of soil moisture (seasonally variable conditions) and soil particle size (seasonally stable conditions).

6. Yield maps

Spatial distribution of plant biomass and other bio-physical parameters within fields is usually not even but very patchy, defined by soil condition, therefore it is possible to compile yield maps. Such maps, created on the basis of satellite images acquired in many seasons represent the spatial variability in crops yield regardless of plant species. Yield maps can also be developed using ground-based methods, but Long et al. (1995) concluded that the remote sensing methods gives much better results. Yield maps are applied to determine patterns of fertilization and irrigation, and indirectly, can be used in planning programs for the eradication of weeds, pests and plant diseases. In such programs, fertilizer or pesticides doses are adopted to the productivity of a specific spot on the field. Apart from economic profits to the farmer also important are environmental benefits that rely on reducing the negative impact of agriculture production inputs on the environment.

7. Soil maps

Soil maps are another type of maps developed using remote sensing data. These maps can be compiled on the basis of airborne or satellite images acquired when the degree of soil coverage by plants is less than 30–50 %. Soil maps present homogenous soil zones with similar properties and conditions for plant growth. These maps are useful in determining soil sampling locations for detailed studies of soil, soil moisture sensors location or developing irrigation plans. Soil maps are often prepared also on the basis of ground measurements of electromagnetic conductivity (EC), which is strongly correlated with the particle size distribution. The satellite data provides information about the properties of soil to a depth of 1–2 cm and with data on EC to a depth of 1 m.

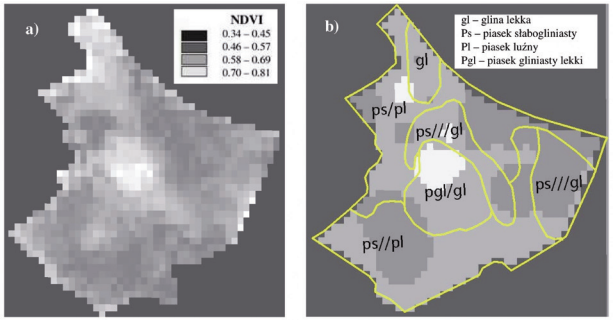


Fig. 3. Variation of the NDVI values calculated for pixels from satellite TERRA Aster image taken on April 24, 2005 for an abandoned fields a) and a soil map put on a classified image b)

Rys. 3. Zmienność wartości wskaźnika NDVI obliczonego dla pikseli obrazu satelitarnego TERRA Aster odłogowanego pola zarejestrowanego 24 kwietnia 2005 a) oraz mapa glebowa nałożona na sklasyfikowany obraz b)

Soil maps can be created using the indirect method by analyzing the spectral variability of vegetation, which grows on this soil. Strong relationship occurring between the spectral data obtained from the satellite or airborne images and the green biomass indicates an indirect connection of these spectral data with soil parameters on fallow fields (Fig. 3). Since the dependence between soil properties and spectral data from the images is of an indirect character this is why a stronger relationship was observed between complex characteristics of the root zone which influenced the fertility of the soil rather than the values of particular soil parameters (Piekarczyk et al. 2011).

8. Variable rate application

Using yield and soil maps, farmers are able to apply fertilizers and herbicides at variable rates within their fields using a Global Positioning System (GPS) receiver mounted for navigation on a sprayer. Usually soil maps obtained on the basis of remote sensing data are used to determine different doses of nitrogen (N) within a field for pre-sowing fertilization (variable rate application). Costs of variable rate N application carried out on the basis of recommendations from the soil map may be about 17% lower than the traditionally used fertilization scheme assuming the same, high N dose across a whole field. N fertilization can be also performed during the growing season when plants cover the soil and on the basis of the airborne or satellite data maps of spatial distribution of NDVI or chlorophyll are compiled. NDVI values are very strongly correlated with chlorophyll and nitrogen content in plants. If the concentration of chlorophyll in plants is too low it means that the plants suffer from a deficit of nutrient and it is necessary to apply additional nitrogen fertilization.

9. Conclusions

A key benefit of the use of remote sensing methods in agriculture is the possibility to obtain in a short time large amounts of data on the condition in which agricultural production takes place. The short review of applications of remote sensing techniques in agriculture presented above leads to the conclusions that in order to fully exploit opportunities offered by these methods a close cooperation of specialists from various fields is required. Their activities should include collecting and processing image data to form of reflectance factors in optical or radar wavelengths, vegetation indices or temperature. In the next step, information about the properties of plants and soils from the processed data are obtained and presented in the form of maps. Then these data, as separate thematic layers, can be integrated with other data (e.g. meteorological) using decision support systems (DSS).

The currently available satellite data are recorded by the sensors with a relatively low spatial and spectral resolution. The results of recent studies indicate that the hyperspectral sensors sampling in relatively narrow portions of the electromagnetic spectrum and identifying unique absorption features of vegetation may allow more accurate extraction of information about specific plant and soil properties than commonly used broadband multispectral satellite sensors. These satellites, collect images with a long revisit time thus the chances of acquiring cloud-free data are relatively small. The problem of cloud cover is less important in the case of airborne photographs, however a difficulty is a calibration of spectral data and georeferencing series of photographs covering large areas. The solution of these problems will be launching into orbits new satellites equipped with more advanced sensors obtaining images with higher spectral, spatial and temporal resolutions.

References

Bochenek Z., Ciołkosz A., Filipiak K. 2000. Szacowanie powierzchni upraw z wykorzystaniem teledetekcji lotniczej i satelitarnej. Fotointerpretacja w geografii – Problemy telegeoinformacji, 31:130–136.

Budde M.E., Rowland J., and Funk C., 2010. Agriculture and Food Availability – Remote Sensing of Agriculture for Food Security Monitoring

in the Developing World. http://www.earthzine.org/2010/02/08/agriculture-and-food-availabilityremote-sensing-of-agriculture-for-food-security-monitoring-in-the-develo ping-world/.

Dąbrowska-Zielińska K., Kogan F., Ciołkosz A., Kowalik W., Gruszczyńska M. 2002. Modelling of crop growth conditions and crop yield in Poland using AVHRR-based indices. International Journal of Remote Sensing, 23(6):1109–1123.

Gates D.M., Keegan H.J., Schleter J.C., Weidner V.R. 1965. Spectral properties of plants. Applied Optics, 4:11–25.

Knipling E.G. 1970. Physical and physiological basis for the reflectance of visible and nearinfrared radiation from vegetation. Remote Sensing of Environment, 1:155–159.

Larsson S. 2002. Acacia canopy cover changes in Rawashda forest reserve, Kassala Province, Eastern Sudan, using linear regression NDVI models. International Journal of Remote Sensing, 23(2):335–339.

Long D.S., Carlson G.R., DeGloria S.D. l995. Quality of field management maps. In: Proc. Site-Specific Mgmt. for Agric. Sys.. 27–30 March 1994, Minneapolis, MN, ASA-CSSA-SSSA, Madison, WI, pp. 251–271.

Miller D., McCarthy J., Zakzeski A. 2009. A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing. Section on Government Statistics – JSM, pp. 3144–3155.

Piekarczyk J., Sulewska H. 2011. Winter oilseed-rape yield estimates from hyperspectral radiometer measurements. Quaestiones Geographicae, 30(1): 19–26.

Piekarczyk J. 2011. Szacowanie plonów roślin uprawnych na podstawie naziemnych pomiarów spektralnych. Teledetekcja Środowiska, 46:23–28.

Piekarczyk J., Kaźmierowski C., Królewicz S. 2011. Relationships between Soil Properties of the Abandoned Fields and spectral Data derived from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). Advances in Space Research, doi: 10.1016/j.asr.2011.09.010.

Scanlon T.M., Albertson J.D., Caylor K.K., Williams C.A. 2002. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. Remote Sensing of Environment, 82:376–388.

Sims D.A. and Gamon J.A. 2003. Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features. Remote Sensing of Environment, 84:526–537.

Tys J., Piekarski W., Jackowska I., Kaczor A., Zając G., Starobrat P. 200. Technologiczne i ekonomiczne uwarunkowania produkcji biopaliwa z rzepaku. Acta Agrophysica, 99:1–153.

GEOINFORMATICA POLONICA
13: 2014

ROKSANA ZARYCHTA[1]
roxana_z@vp.pl
ADRIAN ZARYCHTA[2]
adrian.zarychta@o2.pl

BASIS OF GEOSTATISTICS FOR SCIENTISTS SPECIALISING IN NATURAL SCIENCES BASED ON THE EXAMPLE OF ORDINARY KRIGING

Keywords:

geostatistics, kriging, Walvoort, E{Z}-Kriging

Abstract

The paper presents four stages of acquiring the knowledge about kriging. The second stage has been focused on application of the free software E{Z}-Kriging, which allows to learn the empiric basis of ordinary kriging.

PODSTAWY GEOSTATYSTYKI DLA PRZYRODNIKÓW NA PRZYKŁADZIE KRIGINGU ZWYCZAJNEGO

Słowa kluczowe:

geostatystyka, kriging, Walvoort, E{Z}-Kriging

Abstrakt

W niniejszym artykule zaprezentowano cztery etapy dochodzenia do wiedzy o krigingu. Drugi etap powiązano z darmową aplikacją E{Z}-Kriging, która pozwala zapoznać się z podstawami empirycznymi krigingu zwyczajnego.

1. Introduction

According to Namysłowska-Wilczyńska (2006) "geostatistics deals with regionalised variables – random variables whose location in space or time is important. Those are random variables showing randomness and structure in the behaviour that depends on the position, i.e. location of the sample". This young branch of applied statistics is sometimes called spatial statistics, and its origin dates back to the 1930s, 1940s. The works of D.G. Krieg (1951, 1966, 1976) and G. Matheron (1962) of 1950s and 1960s, which are considered to be a flagship of bibliographic references in the field of geostatistics, are worth mentioning here.

[1] University of Silesia, Faculty of Earth Sciences, Department of Geomorphology, Unit of Karst Geomorphology.
[2] University of Silesia, Faculty of Earth Sciences, Department of Applied Geology, Unit of Geology of Deposits.

Kriging, the name of a well known, often used interpolation techniques, derives from the surname Krieg. It is a method of estimating the value at points of an interpolation net based on the known values at the sampling points. Currently, different types of kriging are known, including: ordinary kriging, simple kriging, probability kriging etc., ordinary point and block kriging are the most common ones applied in natural sciences. The nature of the method itself is quite complicated, since implementation of kriging techniques to the research process requires not only theoretical knowledge of mathematical, statistical sciences, but also experience in its application.

Young researchers who, in the course of research and development, deal with graphical presentation of spatial data in the form of digital terrain models and hypsometric maps are not always familiar with geostatistics. However, it does not mean that they do not use it. It happens that they do not change the default settings offered by the applied software. However, the produced results are not always satisfactory. Therefore, the generated image can be affected by serious errors, which contribute to its distortion, and thus it results in carrying out erroneous deduction made by the investigator. Hence, a question arises, how to perform precise spatial visualization, which allows to obtain an image of reality as much authentic as possible. Apparently, it is possible to use adequate geostatistical methods (including kriging). However, their understanding requires time and experience. Consequently, there is another question, how to understand effectively geostatistical methods (including ordinary kriging), to make their implementation to the analysed example (in order to solve a research problem) the most accurate and justified.

This article aims to provide answers to the above question, by presenting four stages of acquiring the knowledge about ordinary kriging. However, special attention has been paid to the second stage, which includes application of a useful geostatical tool E{Z}-Kriging (which also means that its evaluation has been done), allowing to get acquainted with empiric basis of ordinary kriging.

2. ”Geostatistical exploration”

2.1. Stages of acquiring knowledge about ordinary kriging

Ordinary Kriging is one of the interpolation methods, however, its implementation to the conducted analysis requires large theoretical knowledge and experience in application of geostatistics. Thus, the authors of the paper have distinguished, for the purpose of this article, four stages of acquiring knowledge about the discussed type of kriging (stages may also apply to other methods):

- stage 1: Analysing the theory,
- stage 2: An attempt to understand the theory through experience,
- stage 3: Gaining practical skills as a result of attempts to answer the research problem,
- stage 4: Developing practical skills in the course of iterative use of appropriate algorithms applied to solve the research problem.

Stage 1 should be understood as a cognitive initial stage, in which the researcher gains knowledge about the possibilities of application of kriging techniques while participating in a series of lectures, laboratories, seminars, and also studying the specialized scientific literature, etc., to be able to use the acquired theory in the next stage.

In the stage 2, the researchers implement their knowledge in practice, gaining a theoretical-empirical basis for application of the ordinary kriging.

Stage 3 is devoted to a thorough analysis of the gained theoretical and theoretical-empirical basis. Researchers solve scientific problems, using the selected software, and using the technique of interpolation. However, practical skills are only developed at the next stage – stage 4, as a result of repeated usage of kriging.

Stage 2 is the most important one, because it is responsible for implementation of the theoretical knowledge into practice, and thus gives the opportunity to acquire empirical skills of kriging usage. Improper gaining of the skills usually results in incorrect research conduct in the following two stages. It is worth noting that E{Z}-Kriging, which is presented below is one of the tools helping to implement properly “the studied” theory of ordinary kriging into practice.

2.2. E{Z}-Kriging and its nature

E{Z}-Kriging (*Ordinary Kriging explained*, version 0.2) is a geostatistical freeware application in English, which allows to acquire, in a very comprehensible way, theoretical knowledge about ordinary point kriging, and block kriging. This utility programme was created in 2002 on the initiative of D.J.J Walvoort from the University of Wageningen in the Netherlands (*Laboratory for Soil & Geology*). It can be used without any restrictions, because it has been made available for public use by the author. E{Z}-Kriging is targeted at various users including students, young researchers and others, who deal with generation of digital images (including hypsometric maps, digital terrain models, etc.) by means of application of the ordinary kriging.

Graphic interface of the E{Z}-Kriging application consists of three modules (fig. 1):

1. Data configuration module
2. Semivariance module
3. Kriging module

7 sampling points (xyz) (marked in blue) are presented in tabular form in module 1. They are the basis which was used to carry out estimation of the value of the point "0" (marked in red) located in any place in the Cartesian coordinate system. The investigator obtains information about the distribution of the points (their location can be changed by the user) owing to the installed interactive graphics. Hence, the described module allows to understand in a practical way what the process of estimation is and how it runs (Fig. 1).

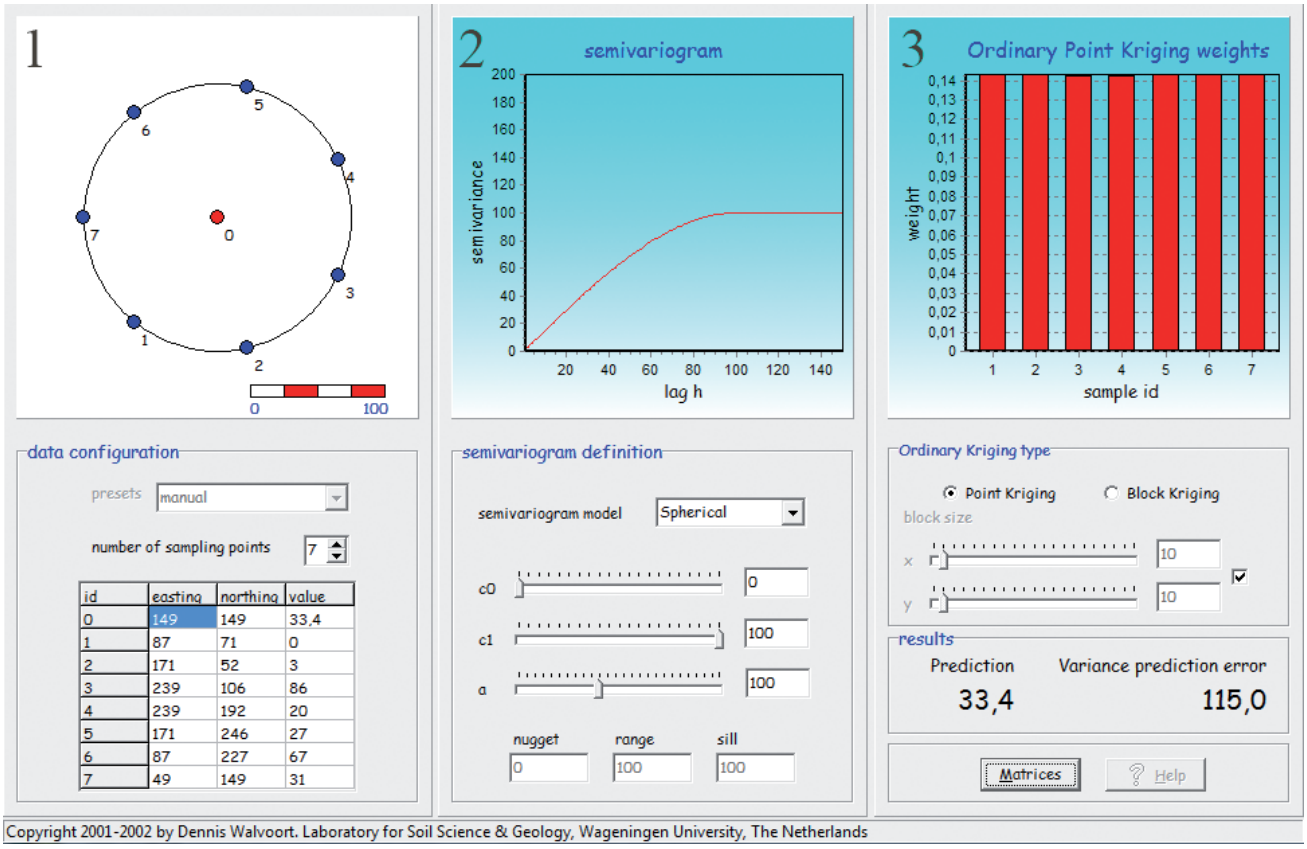


Fig. 1. The graphical interface of E{Z}-Kriging (1, 2, 3 – the modules)
Rys. 1. Interfejs graficzny aplikacji E{Z}-Kriging (1, 2, 3 – moduły)

Module 2 allows the researcher to choose a model of a theoretical semivariogram (out of the three available ones: spherical, exponential and Gaussian), while taking into account the change of its basic parameters (nugget effect, range and sill). Using the module 2 the researcher can observe the semivariogram behaviour with the change of its parameters, thus understands what the structural analysis is (Fig.1).

Module 3 presents the way a matrix is created and clarifies the concept of the condition of unbiasedness of the estimator (where the sum of weights = 1) as well as its importance in ordinary kriging (weight values are presented in a graph in the form of columns). Additionally, it allows to distinguish ordinary point kriging from the block one. As a result, it provides the user with information about the amount of the estimated value at the point "0" and the variance error (Fig. 1, 2).

Considering the above information, it can be stated that the Stage 2 (an attempt to understand the theory through the experience) is important for the researcher, as it is based on the application of E{Z}-Kriging "the researcher gets experience in the ordinary kriging". This is not entirely possible in case of the transition from Stage 1 (analysing the theory) directly to Stage 3 (gaining practical skills as a result of attempts to answer the research problem). It is also prompted by the great potential of geostatistical software in case of application of interpolation techniques (including kriging). Information overload puts an inexperienced researcher in a dilemma when it comes to choosing a proper geostatistical methods and, consequently, contributes to misunderstanding or lack of understanding of its nature.

Until recently, stage 3 has been associated with performing "manual calculations" when applying ordi-

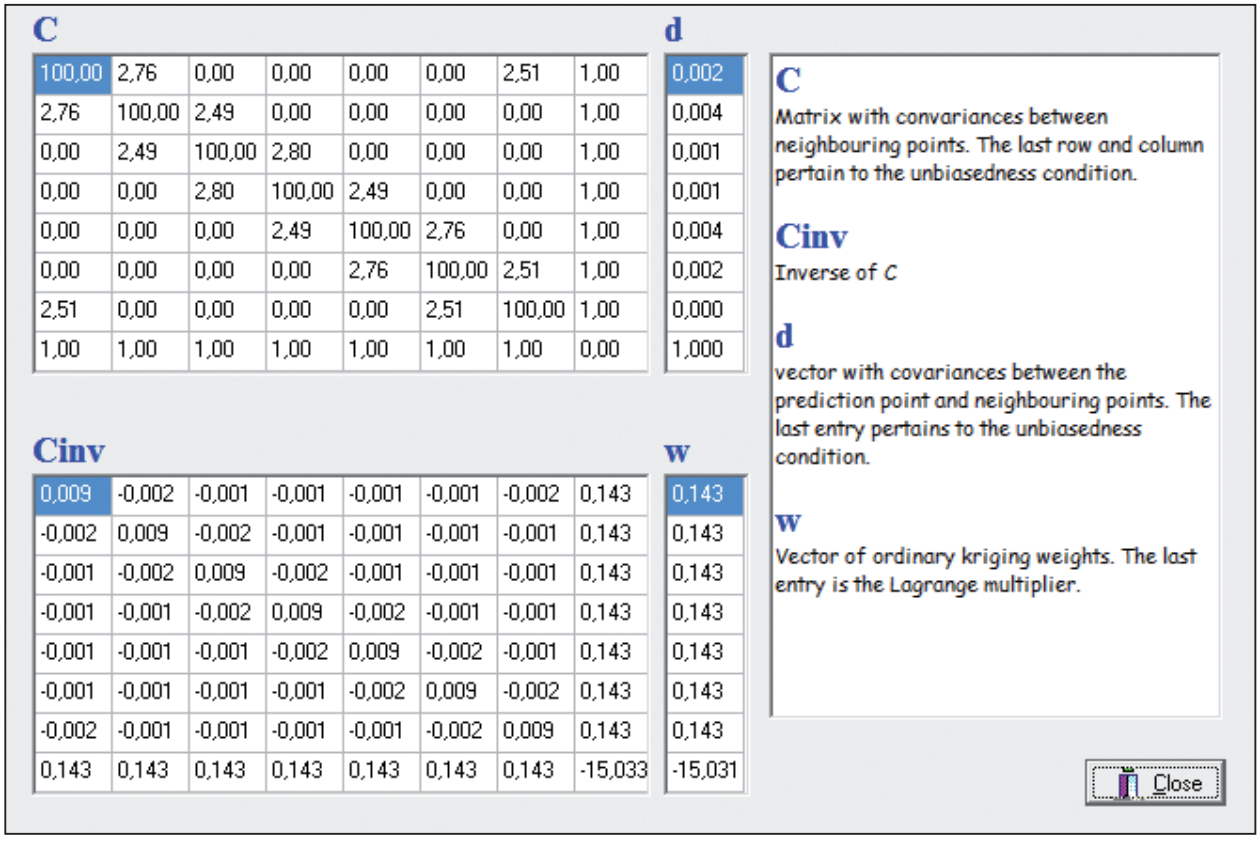


Fig. 2. Data matrices in E{Z}-Kriging
Rys. 2. Macierze danych w E{Z}-Kriging

nary kriging. Currently such complicated and requiring much attention calculations are not necessary.

E{Z}-Kriging application allows monitoring changes (in real time) occurring in the three modules of ordinary kriging due to the interference of the researcher in a given element/parameter (which was not possible in case of performing manual calculations and calculations performed automatically by means of appropriate geostatistical software). For example, a change of the position of a sampling point and/or an estimated point causes changes in the module components. Then other numerical values that build the matrix can be recognised, and owing to that, changed values of the weight and a different value of the estimated point and the error variance can also be distinguished (Fig. 3). A similar situation occurs in case of modifying the semivariogram model, which allows an in-depth empirical analysis of the behaviour of particular parameters of the model.

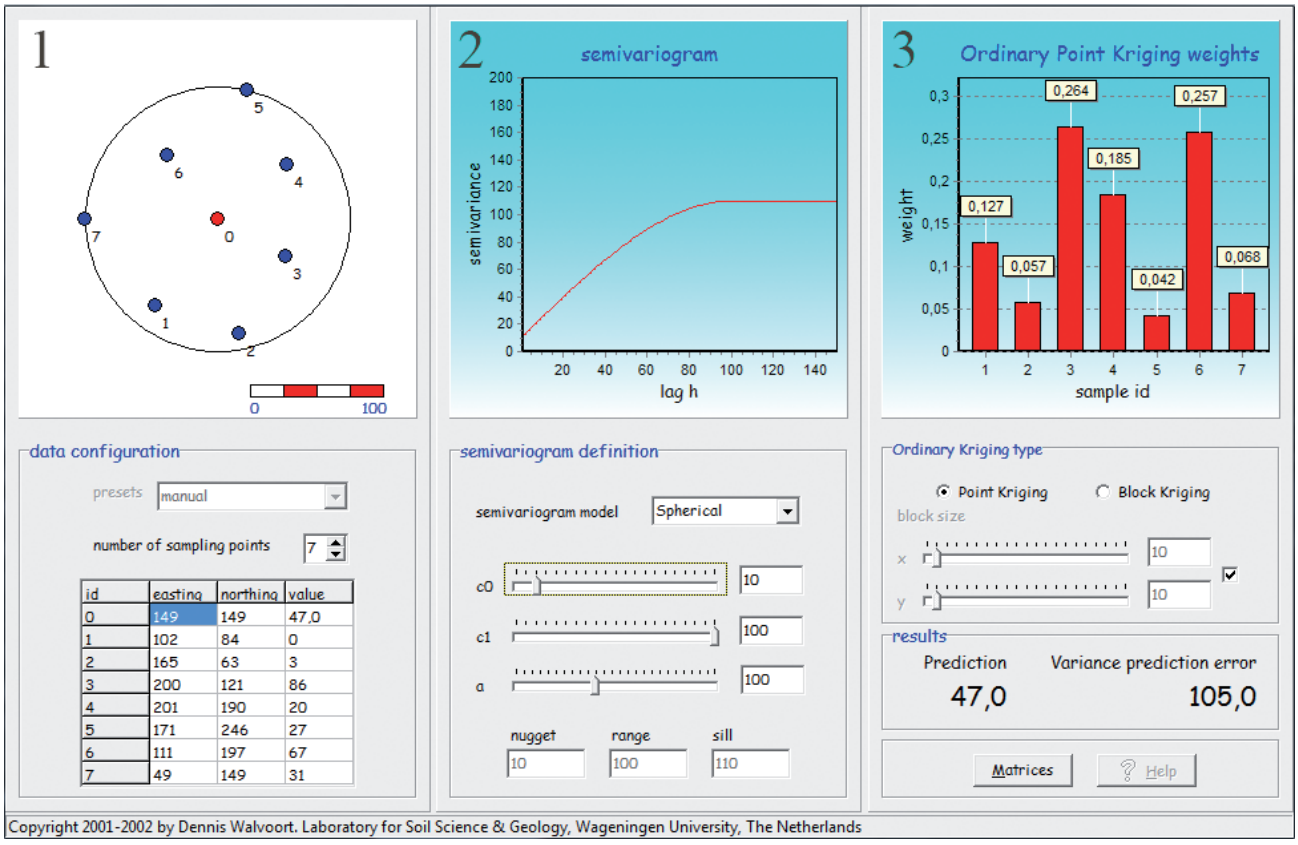


Fig. 3. Change of the location of sampling points and the estimated point and their influence on other modules (1, 2, 3 – the modules)

Rys. 3. Zmiana położenia punktów opróbowań oraz punktu estymowanego i ich wpływ na pozostałe moduły (1, 2, 3 – moduły)

3. Concluding remarks

Ordinary kriging is a relatively complicated geostatical method. It is indicated by a series of mathematical formulas and specialized scientific terminology, which is not always practically justified. Therefore, students and young scientists, before using the kriging method, should acquire theoretical knowledge in geostatistics to gain subsequent "kriging empirical basis". It is possible by means of the freeware E{Z}-Kriging application, which helps potential young researchers involved in the analysis and visualization of spatial data to develop many practical skills. Additional advantages of the program include possibility to observe changes of particular parameters/elements in the three modules included in its graphic interface. However, E{Z}-Kriging also has a number of limitations, namely: it can be applied only to ordinary kriging (it does not apply to other types of kriging such as simple kriging or probability one), it is based on a small number of sampling points and estimation points and it comprises only three theoretical models of semivariograms. Furthermore, it does not solve a research problem (it is possible only when an appropriate geostatical software is applied), because it is only a program presenting the nature of the ordinary kriging.

Nevertheless, E{Z}-Kriging can be used for teaching purposes. It allows to experience and at the same time to understand the basics of the ordinary kriging (nonetheless explaining the issues in dispute and problematic ones concerning the kriging), which are necessary for the correct application of the method when investigating the research problem by means of a geostatical software.

References

Namysłowska-Wilczyńska B. 2006. Geostatystyka. Teoria i zastosowania. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław: 356 pp.

Krige D.G. 1951. A statistical approach to some mine valuation and allied problems at the Witwatersrand. Master's thesis. University of Witwatersrand, Witwatersrand, South Africa: 272 pp.

Krige D.G. 1966. Two-dimensional weighted moving average trend surfaces for ore valuation. Proc. of the Symp. on Mathematical Statistics and Computer Applications in Ore Valuation, Johannesburg, South African Institute of Mining and Metallurgy, Johannesburg: 13–38.

Krige D.G. 1976. Some basic consideration in the application of geostatistics to the valuation of ore in South African gold mines. Journal of the South African Institute of Mining and Metallurgy: 383–391.

Matheron G. 1962. Traité de Géostatistique appliquée, Tome 1. Éditions Technip, Paris: 334 pp.

GEOINFORMATICA POLONICA
13: 2014

MONIKA ŁÓJ[1]
mloj@agh.edu.pl
TOMISŁAW GOŁĘBIOWSKI[2]
goleb@wis.pk.edu.pl
SŁAWOMIR PORZUCEK[3]
porzucek@agh.edu.pl

GEOPHYSICAL SURVEYS AND MODELLING FOR RECOGNIZING OF GYPSUM KARST

Key words:

GPR, microgravimetry, gypsum karst, modelling

Abstract

Geophysical surveys are often used for examination of karst phenomena in limestone. This paper presents the research on gypsum karst which occurrence is much rarer. Non-invasive terrain surveys were carried out over known karst cavity in order to visualize karst phenomena developed around the void. Two geophysical methods that seems to be predestined for karst examination were selected, i.e. microgravimetry and georadar. For the interpretation of measured data numerical modelling was applied. Geophysical interpenetration depicted the presence of weathered material in the near surface zone and fractures and loose zones located over the karst cavity.

BADANIA GEOFIZYCZNE I MODELOWANIE DLA ROZPOZNANIA KRASU GIPSOWEGO

Słowa kluczowe:

Georadar, mikrograwimetria, kras gipsowy, modelowanie

Abstrakt

Pomiary geofizyczne są często stosowane do badania krasu wapiennego. W artykule przedstawiono natomiast opis badań krasu gipsowego, którego występowanie jest o wiele rzadsze. Nieinwazyjne badania terenowe przeprowadzono nad znaną jaskinią krasową w celu zobrazowania zjawisk krasowych rozwijających się wokół pustki. Do badań wybrano dwie predystynowane do tego celu metody geofizyczne: mikrograwimetryczną i georadarową. Interpretację wyników badań geofizycznych przeprowadzono z wykorzystaniem modelowania numerycznego. Wyniki interpretacji wskazują, że w strefie przypowierzchniowej występuje warstwa zwietrzała, a nad pustką krasową występują strefy spękań i rozluźnień.

[1] AGH University of Science and Technology in Cracow, Faculty of Geology, Geophysics and Environmental Protection, Department of Geophysics.
[2] Cracow University of Technology, Faculty of Environmental Engineering, Institute of Geotechnics.
[3] AGH University of Science and Technology in Cracow, Faculty of Geology, Geophysics and Environmental Protection, Department of Geophysics.

1. Introduction

Geophysical methods have been widely used for the exploration and research of karst voids for many years. The most common methods are: microgravimetry and georadar (Bishop et al. 1997; Chamberlain et al. 2000; Rybakov et al. 2001; Mochales et al. 2008; Leucci and Giorgi 2010). However, almost all research were carried out for detection of karst voids formed in limestone. In the context of this publication, the authors chose to research a void created in outcrops of Tertiary gypsum, which is localized in the Wiślica Caves area (Poland). There are a lot of caves created in rod-like gypsum crystals and sabre-like gypsum crystal layers, as a result of karst corrosion at the groundwater level (Flis 1954).

The studied cave, called the Cave in Siesławice (Gubała et al. 1998), is bell-shaped and partially filled with water. Information from 1998 shows that the cave was about 24 metres long, and its ceiling was approximately 1.5÷2 m BGL (Gubała et al. 1998), but measurements taken in 2013 shown that its length is about 17 metres. At present the entrance to the cave is located inside the sinkhole (Figs. 1 and 2), so it can be assumed that this sinkhole was created between 1998 and 2013.

To recognize the rock mass in the area of the Cave in Siesławice two geophysical methods were selected: microgravimetry and georadar. Both methods are very useful for the detection of cavities (BSI 1999, A.S.T.M. 1999). The survey was carried out in two profiles intersecting the cave (Fig. 1). Due to the escarpment, GPR profiles were shorter than the microgravimetric ones.

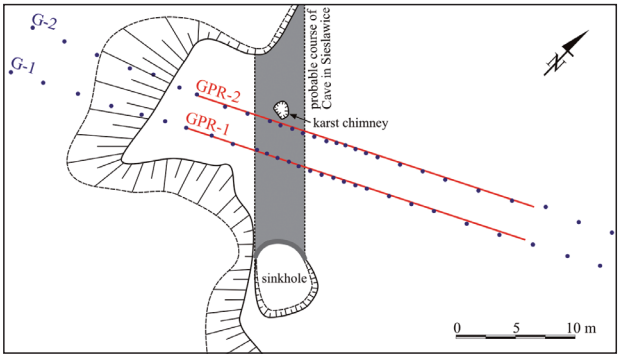


Fig. 1. Localization of geophysical surveys
Rys. 1. Lokalizacja badań geofizycznych

The aim of geophysical investigations was analysis of applicability of selected techniques for recognizing of karst phenomena developed in gypsum. For terrain surveys region over the known void was selected because it was highly probability that around this cavity karst processes should be observed. Karst phenomena may develop both limestone and gypsum rocks but due to the fact that gypsum is weaker than limestone stronger weathering and washing out of material should be expected in gypsum; therefore the results of geophysical surveys carried out for gypsum examination might be different from those recorded during limestone investigation. As it was shown in the paper near surface part of gypsum was weathered and superposition of weathered zone and fractures and loose zones origin from karst processes was observed over the cavity; this superposition of effects influenced to GPR data. For analysis of this influence numerical modelling was applied. Microgravimetric methods is used to find places of different bulk density in the rock mass – it allows to detect voids and loose zones. When the geometrical parameters of the void is known it is possible to calculate its gravity effect and compare with anomaly distribution. It allows to find the loose zones near the void.

Fig. 2. The entrance to the cave – photo taken from inside the sinkhole
Rys. 2. Wejście do jaskini z wnętrza zapadliska

2. Geological setting

The Nida Gypsum deposits have been divided into two cave sub-regions in terms of geological and morphological division. The limit of these sub-regions is the Pilica river. The Cave in Siesławice is located north-west of the Pilica river, in the cave sub-region of Busko – Staszów (Nowak 1986a, 1986b).

The shape of the gypsum basin was formed in the phase of the Alpine orogeny together with the tectonic movement of the Carpathians mountains (Rutkowski 1986). In the geological profile of the basin, the bedrock underlines Upper Jurassic limestone. Above them there are several hundred metres of the Cretaceous layer, developed in the form of lime, clay, silica and marl deposits. The south-eastern part of the Nida Gypsum deposits is covered by tertiary deposits, mainly of the Miocene age, developed in different ways depending on the sedimentation case. Some part of the Tertiary sediments are covered by Quaternary age deposits, which in the southern part are developed in the leoss form (Rutkowski 1986; Walaszczyk 1992).

In the Nida Gypsum deposits area, Miocene deposits developed in three series of sedimentary: gypsum bandage layers, a layer of evaporates (gypsum) and the overgypsum layers (Rutkowski 1986). Gypsum bandage layers are developed in the form of deposits of marl with clay and sand inserts. In these layers biogenic limestone also occurs to replace these deposits.

The evaporate layer thickness is about 40 m and has a different structure and texture (Kasprzyk 1991, 1993). At the base there is glassy gypsum (3.5 m thickness) with giant gypsum intergrowths. Above there is a 2–3 m thickness, row of grass-like gypsum with inter beds of stromatolitic and alabastrine gipsum. The next part of the evaporate deposit is over a dozen layers of skeletal gypsum and sabre gypsum. The last evaporate layer is laminated and detrital gypsum.

The structure of the overgypsum layer depends on the sedimentary zone. In the coastal zone of the basin this layer is developed in clastic sediments and detrital limestone formed with little thickness. In the south-east part of gypsum basin it is developed in clays of increasing thickness (Łaptuś 1992).

The Cave in Siesławice is located in sabre gypsum layer (Fig. 2). The karst form develops in all parts of the evaporate series (Turchinov 1997), until about 70% appears in skeletal gypsum and sabre gypsum. Due to the way it develops the layer described above – several centimeters of gypsum bank and their not too high mechanical strength resulting in collapsing – the karst void is usually a bells or a lenses shape. In the glassy gypsum layer the karst void develops mainly across the fractures. It has got a trapezoidal shape and rarely collapses (Wołoszyn 1990).

3. Geodetic measurements

An indispensable part of microgravity research are geodetic measurements. They are used to determine the XYZ coordinates of each point. Especially important is the accurate measurement of height, which is needed to calculate the necessary gravity corrections: free-air and Bouguer correction. Surveying was done using an electronic total station, and thus maintained the accuracy of all three orthogonal systems within 0.01 m.

Also the characteristic elements of the terrain were surveyed, allowing the reflection of the terrain around the test cave. Unfortunately measurements of the full shape of the cave could not be carried out, because it was flooded to a depth of about 1 m (Fig. 2). Of course, the water level is dependent on the season. Therefore the shape of the entrance to the cavity was measured and it was used for further analysis.

The cave has two small holes on the opposite side to the entrance. Their location has allowed the determination of the probable course of the cave (Fig. 2).

4. Principles of microgravimetric method

The gravimetric method is the study of variation in the gravitational field (gravity field) on the surface of the Earth (Telford et al. 2004). The variability of this field is related to variation of bulk density, which is accorded to the kind of rock mass building geological environment. Any change in the value of the density generates a change in the distribution of the gravity field.

A variation of the gravimetric method is the microgravity method. It enables tracking of shallow geological structures and geotechnical zones, as well as identification of shallow underlying voids or loosening of the rock mass (Sharma 2004; Porzucek 2013).

A gravity survey was carried out using a CG-5 Autograv gravimeter by the Canadian company Scintrex, whose measurement accuracy is 0.005 mGal, with repeatability of 0.001 mGal. At each point the measurement was taken twice to avoid random error. The survey was considered correct when the measured value of the gravity differed by less than 0.005 mGal. Otherwise, the measurement continued until at least three different values of less than 0.005 mGal were obtained. An important element of the survey was accurate elimination of drift, which was obtained with reference to the base point at approximately 1 hour.

Data processing was performed using the programs included in Geosoft's Oasis Montaj package.

Microgravity research was performed on two parallel profiles spaced apart by approximately 2.75 m (Fig. 1) with a length of about 53 m each. Measuring points were located every 2 m, and in accessible places. Directly over the cave, measuring points were located every 1 m.

In the first stage of data processing, Bouguer anomaly values were calculated. The calculations were made on the basis of a generally known formula (1):

$$\Delta g_B = g_p + 0{,}3086 \cdot \Delta h - 0{,}04193 \cdot \rho \cdot \Delta h + \Delta g_t - \Delta\gamma \quad (1)$$

Δg_B – Bouguer anomaly; mGal,
g_p – survey value of microgravity; mGal,
Δh – elevation; m,
ρ – average bulk density of gypsum; 10^3 kg × m^{-3},
Δg_t – terrain correction; mGal,
$\Delta\gamma$ – normal value of gravity; mGal

Both measuring profiles were carried out near the slope, and additionally, the sinkhole was located close to them, which is associated with the studied cave (Figs. and 2). Shaping the land surface requires the calculation of the terrain corrections. Based on previous experience (Porzucek 2013), additional surveying was carried out, so that at each point it was possible to obtain information about the terrain of a radius of 25 m. For the calculation of the terrain corrections and Bouguer corrections assumed average gypsum density was equal to $2.3 \cdot 10^3$ kg·m^{-3}. The terrain correction was calculated using the Geosoft Oasis Montaj package.

The value of the normal gravity was calculated based on a formula specified by Blakely (1991).

5. Principles of georadar method

The georadar method, also called GPR (i.e. Ground-Penetrating Radar), is a non-invasive geophysical technique that uses electromagnetic waves in the range of 10 MHz to a few GHz for examination of geological media as well as anthropogenic constructions. The most popular mode of GPR survey is the so-called constant-offset (zero-offset might be assumed) reflection profiling mode, which is briefly described below.

GPR surveys in constant-offset refection profiling mode are similar to scanning. During measurements two antennae, i.e. a transmitter antenna (Tx) and receiver antenna (Rx), move along the profile (Fig. 3) and the transmitter antenna emits electromagnetic signals with an assumed distance interval (Δx). The electromagnetic signals

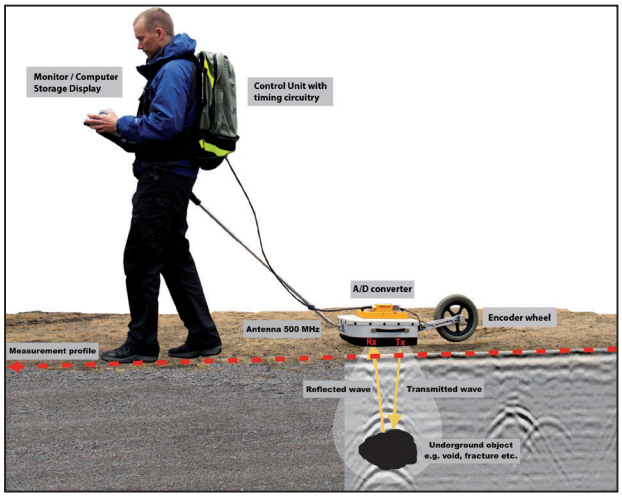


Fig. 3. GPR equipment and survey in constant-offset reflection profiling mode (www.malags.se)
Rys. 3. Aparatura georadarowa i pomiar w standardzie stałooffsetowego profilowania refleksyjnego

propagate in the examined medium as electromagnetic waves and the reflected waves are recorded by the receiver antenna (Fig. 3). As a result, a digital radargram with a distribution of reflections in a distance-time plane is obtained. During data processing the vertical time axis (in nanoseconds) is converted into the depth axis (in metres).

Propagation of electromagnetic waves depends on the material properties of the examined medium, i.e. relative electrical permittivity ε_r [-] and conductivity σ [mS/m]; the third material parameter, i.e. relative magnetic permittivity μ_r [-], is not taken into account during analysis of GPR data and is assumed as 1.

In reflection profiling mode the reflection of electromagnetic waves from underground objects and/or geological boundaries is described by the reflection coefficient R [-]:

$$R = \frac{\sqrt{\varepsilon_{r\_host\_medium}} - \sqrt{\varepsilon_{r\_object}}}{\sqrt{\varepsilon_{r\_host\_medium}} + \sqrt{\varepsilon_{r\_object}}} \qquad (2)$$

For a lossless medium like gypsum, i.e. for a medium with negligible electrical conductivity, the following simplified formulae, describing velocity v [m/ns] and ohmic attenuation α_{ohmic} [dB/m], might be assumed:

$$v = \frac{0.3}{\sqrt{\varepsilon_r}} \qquad \alpha_{ohmic} = 1.7 \cdot 10^3 \frac{\sigma}{\sqrt{\varepsilon_r}} \qquad (3, 4)$$

In the GPR method, depending on the analyzed problem, antennae with different constructions, frequencies and polarizations are used. A description of the antennae used for GPR surveys at the Siesławice site will be presented later in the paper.

6. Microgravity surveys and interpretation with modelling

After processing the microgravimetric data Bouguer anomaly values were obtained (Δg_B), whose distributions are shown in Fig. 4. On both profiles 0m was taken at the beginning of the GPR profiles. On the curves minor disturbances are visible coming from the random errors or the gravitational effects of very shallow bodies. For their elimination a low pass Fast-Fourier Transform Butterworth filter with a central wavelength cut-off of 5 m was applied to give smoothed Bouguer anomaly distributions Δg_{BS} (Fig. 4).

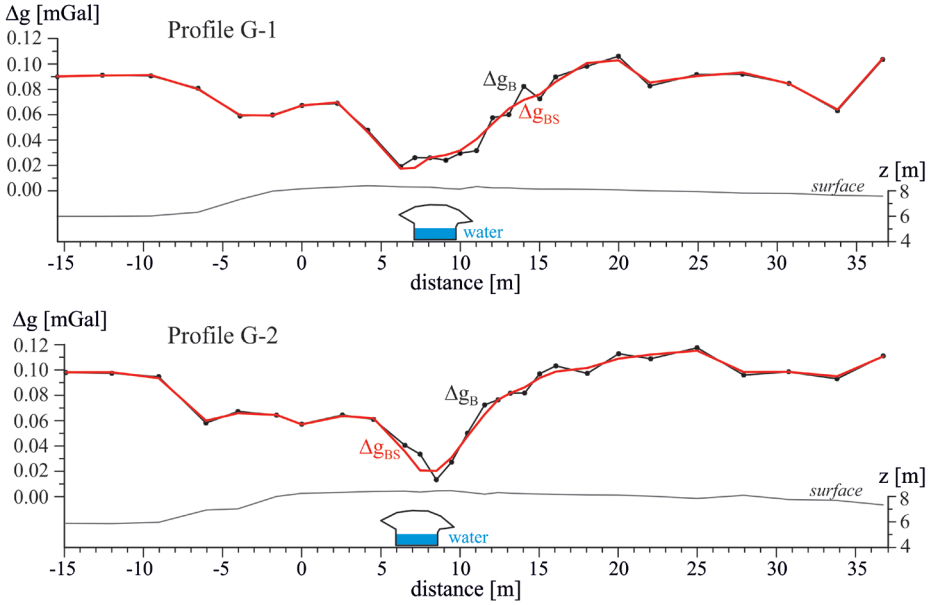


Fig. 4. Distributions of Bouguer anomalies before and after smoothing
Rys. 4. Rozkłady anomalii Bouguera przed i po wygładzeniu

Both profiles of the Bouguer anomaly distributions (Δg_{BS}) clearly indicate the gravitational influence of the cave, which is visible as a relatively negative anomaly over the void. As is known, the measured values of gravity are the sum of two components: the regional field and the local field. Analysing both distributions the small impact of the regional field can be seen. In other words, the values of gravity, disregarding the effect of local anomalies, increase slightly on the profiles.

To eliminate the regional component, a filtration in the frequency domain and a polynomial approximation of the regional field were applied. Determining regional component is complicated by the fact that its amplitude is much smaller than the amplitude of the strong local anomalies. Filtration was calculated in the frequency domain by Butterworth filter. Fig. 5 shows the results of filtration for the central wavenumber of the filter which equal 30 m, 50 m, 75 m and 100 m (Δg_{R_B30}, Δg_{R_B50}, Δg_{R_B75}, Δg_{R_B100}). For a polynomial approximation of a regional field a linear polynomial was selected, and the calculation was made in two versions. In the first version of the calculation all stations on the profile were used (Δg_{R_P1a}), and in the second version the subjective stations were chosen, in which the impact of local anomalies is not visible (Δg_{R_P1s}).

Analysing the distributions, obtained by filtering in the frequency domain, on both profiles the calculated regional anomaly is devoid of the influence of local anomalies only for wavelengths equal to 100 m. Distribution of regional anomalies for this wavelength practically corresponds to a straight line, whose values increased on the profiles. The slope of this line and lines obtained by approximation of a regional field by line polynomial, for both versions, are very similar. Based on the regional field the residual anomalies were calculated, which are a mathematical reflection of local anomalies. Since the interpretation is based on the relative gravity values, the values of residual anomaly profiles were subjected to a specific normalization. For each the calculated instance minimum value was found, then the anomalies were re-

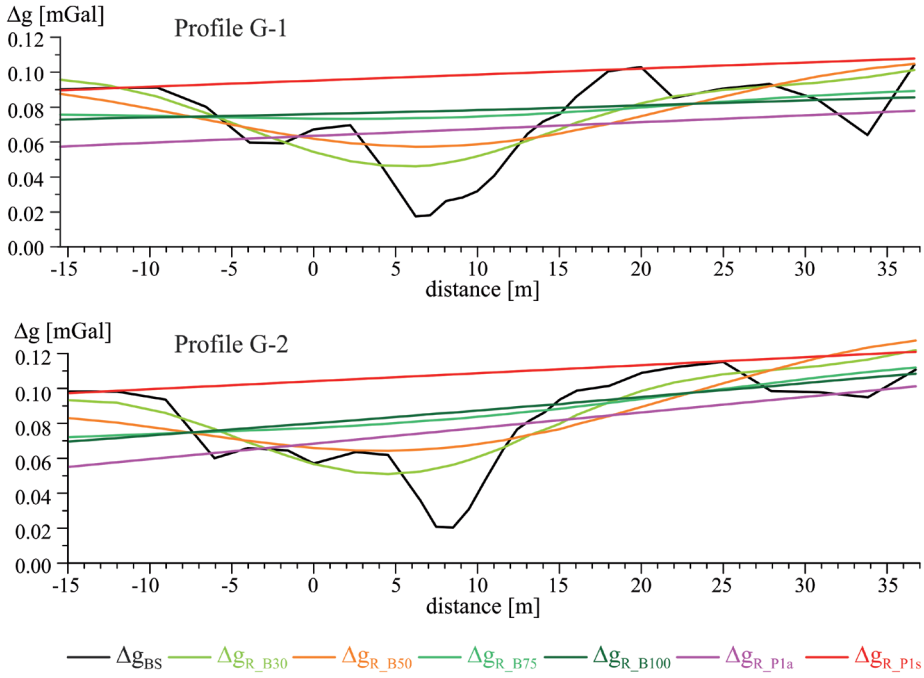


Fig. 5. Distribution of regional anomalies calculated by different methods
Rys. 5. Rozkłady anomalii regionalnych wyliczonych różnymi metodami

calculated by subtracting this minimum, which meant that the minimum value for each case was zero. In the final stage the minimum minimorum of the initial calculations was subtracted from each value. With such normalization the minimum value for each profile was the same. The results of the normalization are illustrated in Fig. 6.

In profile G-1, distributions of residual anomalies, calculated by various methods, are almost identical. In the second profile, G-2, distributions are identical in the region of the cave, but differ slightly beyond its gravitational influence. Thus, for further quantitative interpretation it does not matter which method is selected and it is almost certain that separated the residual anomalies are correct. Authors selected the residual anomalies obtained after approximation of a regional field by linear polynomial using values from all stations (Δg_{R_P1a}).

In the next stage of interpretation, based on the measured shape of the cave, its gravitational effect was calculated using 2.5-D modelling of the GM-SYS Profile Modelling, which is part of the Oasis Montaj package. The modelling results are shown in Fig. 7, together with the distributions of the residual anomalies.

Analysing the distributions on both profiles (Fig. 6) is clearly visible the correlation between the gravity effect of the model of the cave and residual anomalies, the origin of which should be associated with the existence of the void. However, there are also differences between the compared distributions.

In profile G-1 the model anomaly has a similar amplitude to the amplitude of the residual anomaly, but far too little horizontal range. There may be two reasons for this: the other dimensions of the cave and/or the existence of loosened and small voids in the gypsum surrounding the cave. Distributions in the second profile, G-2, indicate, however, that the axis of the cave should be moved by about 2 m to the right, and also that the ceiling of the cave is closer to the surface or there are cracks over it.

On both profiles a decrease in values of the residual anomaly to the left of the cave is clearly visible (Fig. 6). Its location indicates that it should be associated

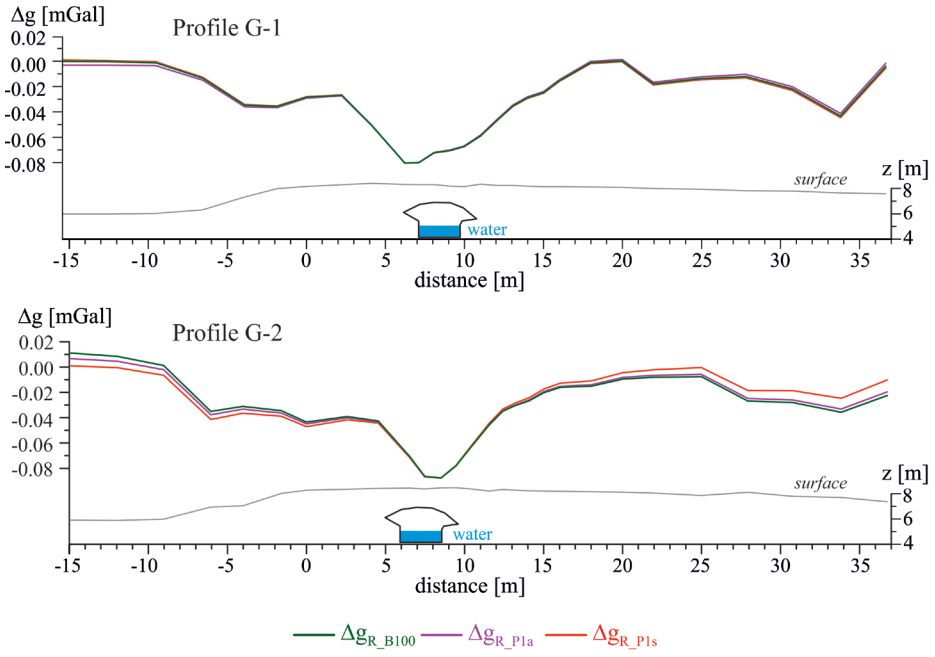


Fig. 6. Distributions of residual anomalies calculated by selected methods
Rys. 6. Rozkłady anomalii rezydualnych wyliczonych wybranymi metodami

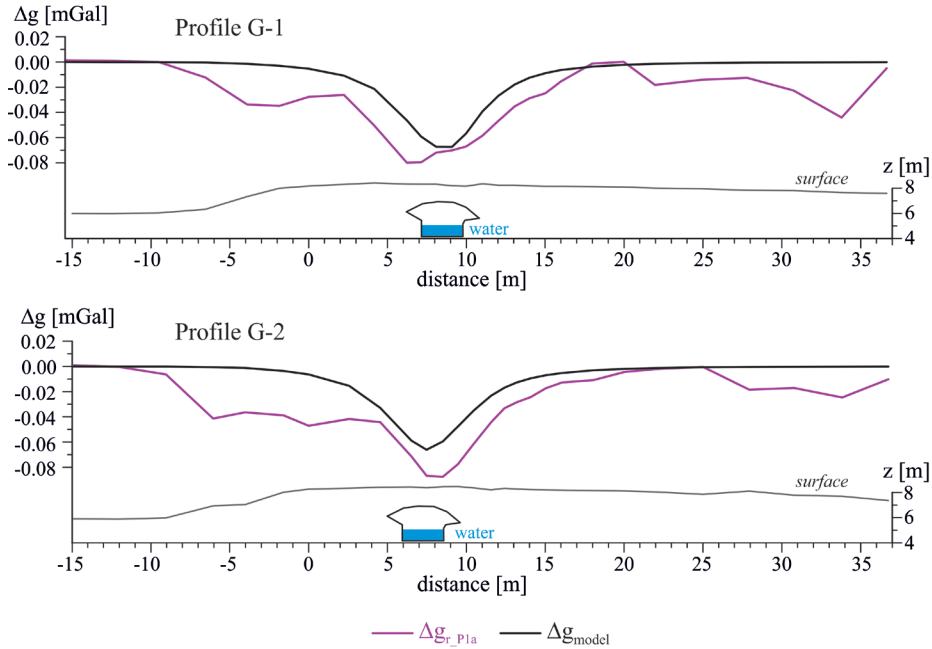


Fig. 7. Distributions of residual anomalies and the gravity effect on the model of the cave
Rys. 7. Rozkłady anomalii rezydualnych i efektu grawitacyjnego od modelu jaskini

with a decrease in the average density of gypsum rock on a slope and its direct surroundings.

On the distribution of the residual anomalies in profile G-1, ranging from 20 m, a lowering of the value is seen. This correlates with a similar decrease in the profile of G-2. It should be assumed that in this region there are few subsurface voids or loosenings.

7. Georadar surveys and interpretation with modelling

At the Siesławice site GPR surveys were carried out for preliminary recognition of the near surface zone, so terrain measurements were conducted only in 2D mode, along two profiles (Fig. 1), in reflection profiling mode. Traces were recorded along profiles with distance internal $\Delta x = 0.025$m. For improvement of the signal/noise ratio, stacking 8 times was defined during data acquisition.

GPR measurements were carried out using the ProEx system (MALA Geoscience, Sweden) with shielded antennae with a frequency of 500 MHz (Fig. 3). Polarization (orientation) of Tx and Rx was perpendicular to the profile line, max. depth penetration was c.a. 5 m and mean resolution was equal 0.05m.

All radargrams presented in the paper are shown in normalized form, i.e. with amplitudes of reflections normalized to the max. amplitude of the direct air wave. For time-depth conversion a constant mean velocity equal to 0.11 m/ns was assumed. Such velocity is adequate to the results of laboratory tests where values of the relative dielectric permittivity of gypsum were determined in the range from 4 to 7 (Table 1) and consequently velocities of electromagnetic wave changed from 0.11 to 0.15 m/ns. Assumed velocity allowed approximated positioning of reflections from the roof of the cavity in profile GPR-2 (Fig. 8B). The distribution of reflections from the roof of the cavity (anomaly "B" in Fig. 8B) depicts that its shape is differ from that measured by the entrance to the cave. Determination of reflections from the roof of the cavity in profile GPR-1 (Fig. 8A) is not possible due to masking effect generated by fractures and loose zones located over the cavity.

Radargrams were digitally processed with application of advanced processing and the following procedures were used: t_0 and topographic corrections, removal of wowing effect, DC correction, Butterworth time-dependent filtration, 1D median filtering, gaining with the use of energy decay function, background removal, stacking, 2D average, spectral whitening, morphologic filtration. Detailed description of applied procedures as well as description of parameters assumed for the mentioned procedures can be found in books by Annan (1999), Golebiowski (2012), Karczewski (2007), ReflexW Manual (2013). Radargrams after processing are depicted in Fig. 8.

In Fig. 8, the near surface weathered zone (WZ), with a mean thickness of c.a. 1.5m, can easily be recognized. Presence of weathered zone and especially superposition of weathered zone and fractures and loose zones developed over the cavity due to karst processes has a strong influence on the reflections recorded at the greater depths. Masking of reflections from the cavity was the result of scattering attenuation of signals on the fractures and loose zones which accrued in the weathered zone. High amplitudes of reflections in the weathered zone were caused by the high value of the reflection coefficient R on the boundaries between fractures and loose zones filled with air (|R|=0.35) or water (|R|=0.6) and solid gypsum.

For comparison, in Fig. 9A a radargram is shown recorded over similar underground objects, i.e. over shallowly located tunnels (Fig. 9B), and diffraction hyperbolae from tunnels are clearly visible in Fig. 9A. When the weathered zone does not occur or has a small thickness and fractures and loose zones are not strongly developed over the void reflections from the roof of the void are easy to interpret (Fig. 8B, Fig. 9A). Unambiguity distinguishing of reflections from the roof of the cavity in Fig. 8A is not possible due to superposition of weathered zone and fractures and loose zones developed over the void in karst processes; this problem will be analyzed in the further part of the paper.

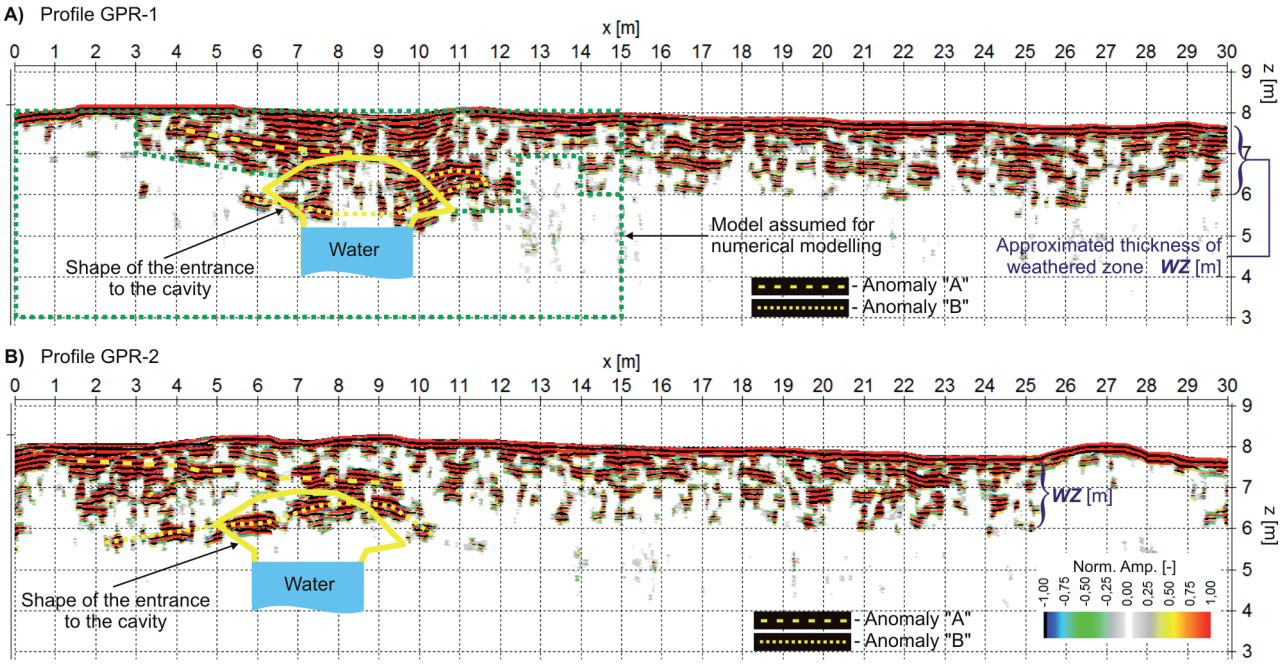


Fig. 8. Radargrams for profiles GPR-1 (A) and GPR-2 (B)
Rys. 8. Echogramy dla profili GPR-1 (A) i GPR-2 (B)

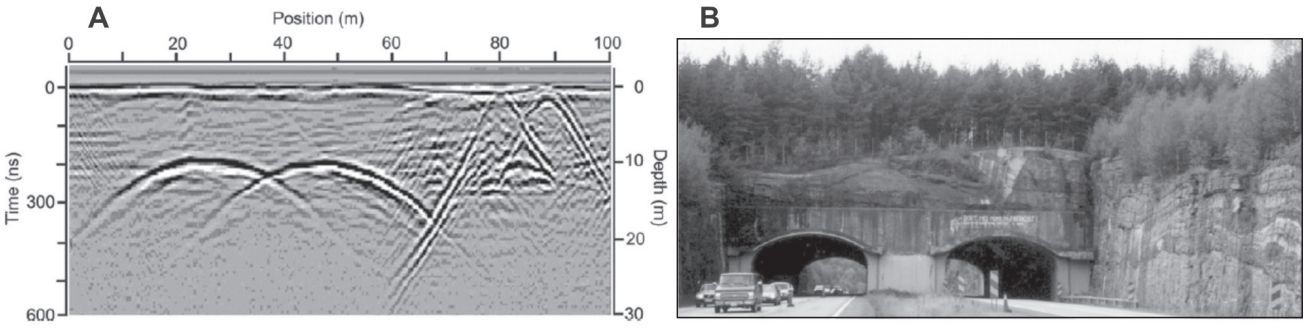


Fig. 9. Radargram (A) recorded over tunnels (B) in lossless medium (Jol 2009)
Rys. 9. Echogram (A) zarejestrowany nad tunelami (B) w ośrodku bezstratnym

In deeper parts of the rock mass, beside the cavity region, only a few low-amplitude anomalies can be seen, which deliver information about the good consolidation of gypsum and low level of heterogeneity. Lack of reflections in deeper parts of the rock mass might also be caused by higher attenuation of signals in clayey gypsum. Analysis of samples taken from the cavity allows to assume that amount of clay minerals in gypsum should not affect strongly to signals attenuation.

In Fig. 8A reflections from the floor of the cavity (covered by water) are visible; their position at lower depth than the real position of the floor (determined from geotactic measurements) was caused by higher velocity of electromagnetic wave in air inside the void (v = 0.3 m/ns) than velocity assumed for time-depth conversion, i.e. v = 0.11 m/ns. Also shape of the floor on the radargram (Fig. 8A – anomaly "B") is differ from that measured by the entrance to the cave; this fact depicts that shape of the cavity is varied in space.

In both radargrams (Fig. 8) reflections from any geological boundary or horizontal crack or anastomosis are clearly visible (anomaly "A").

The aim of numerical modelling was the analysis of the masking effect generated by the weathered zone and in consequence by scattering attenuation. Because masking effect is observed only in Fig. 8A, therefore modelling was carried out only for profile GPR-1.

Numerical models were built on the basis of information from Fig. 8A and the following situations were analyzed: Fig. 10A – model without weathered zone, Fig. 10C – model with weathered zone filled with air, and Fig. 10E – model with weathered zone filled with fresh water.

Material properties were introduced into the models on the basis of information presented in Table 1. Due to negligible change of terrain morphology (Fig. 8A) upper boundary of numerical models was assumed as a flat boundary.

Modelling was performed with the use of the option Modelling of ReflexW program, which is based on the FDTD (Finite Difference Time Domain) technique (Yee 1966). A detailed description of the application of the FDTD technique for GPR modelling can be found, among others, in papers by Golebiowski (2004, 2007), therefore theoretical considerations will be omitted in this chapter.

The weathered zone composed of randomly distributed fractures and loose zones can be characterized by violent changes in electromagnetic properties (ε_r and σ) in respect of the solid medium. This results from the fact that fractures have the ability to accumulate different media; in the cases analyzed in this paper it is air and fresh water (from precipitation and melted snow). Because there is no possibility of presenting a detailed spatial distribution of fractures and loose zones in the weathered zone, only a stochastical model of this zone can be prepared which was generally described below; detailed description of stochastical model using in numerical analysis of GPR data can be found in books by Golebiowski (2012) and the ReflexW Manual (2013).

In the first step of weathered zone construction in the numerical model, "i" and "j" coordinates of first grid

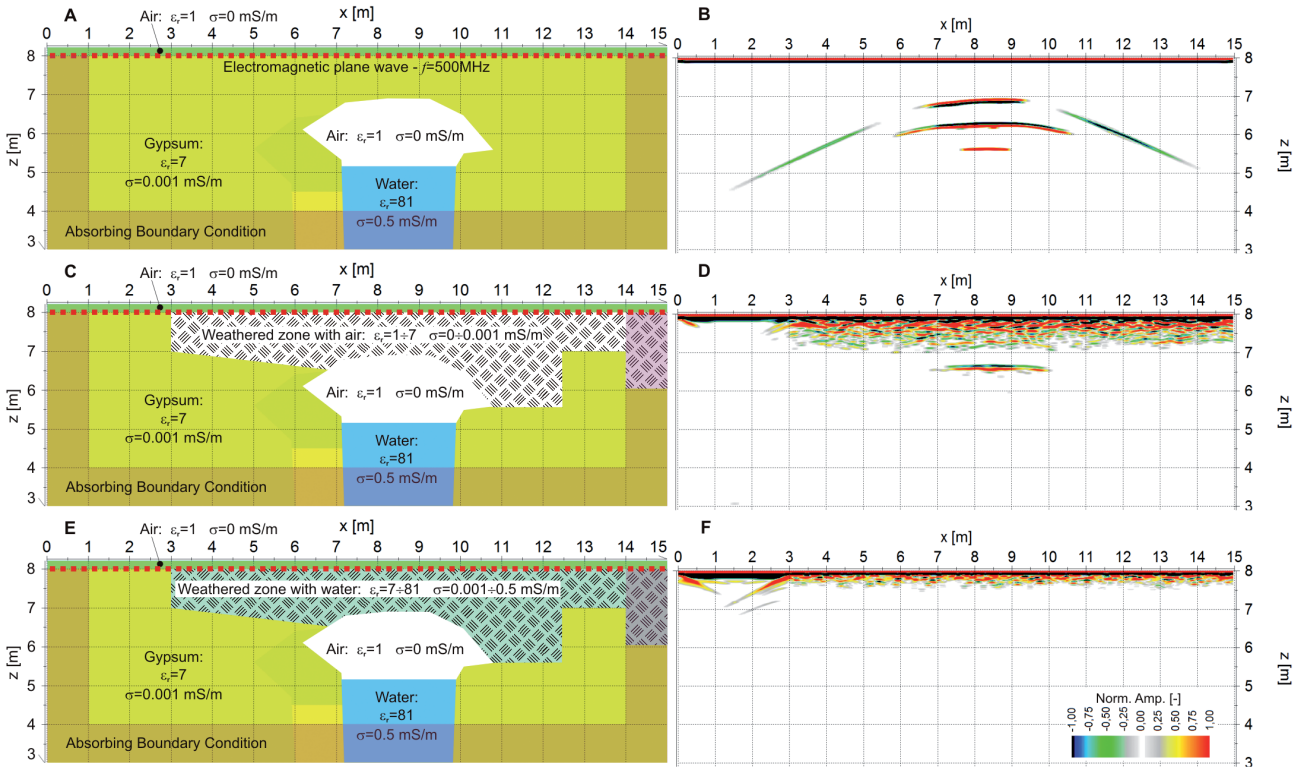


Fig. 10. A) Model without weathered zone; B) Synthetic radargram for model A; C) Model with weathered zone filled with air; D) Synthetic radargram for model C; E) Model with weathered zone filled with water; F) Synthetic radargram for model E

Rys. 10. A) Model bez strefy zwietrzałej; B) Echogram syntetyczny dla modelu A; C) Model ze strefą zwietrzałą wypełnioną powietrzem; D) Echogram syntetyczny dla modelu C; E) Model ze strefą zwietrzałą wypełnioną wodą; F) Echogram syntetyczny dla modelu E

point are randomly chosen; afterwards for chosen grid point a random generator prescribes values of ε_r and σ from the range between properties of gypsum and water or gypsum and air (Table 1); in such way chosen grid point represents either gypsum or water or air or equivalent properties of mixture of mentioned media; equivalent properties describe different saturation of porous space of gypsum by water and/or air. In the next step second grid point belonging to the weathered zone is randomly chosen and etc. The scheme presented above was repeated as long as all grid points in the weathered zone were filled with randomly determined values of ε_r and σ.

The source of the electromagnetic wave in the models was described as a plane wave (Fig. 10A, C, E) and the source function was defined on the basis of discretization of a real 500 MHz signal, taken from the measured radargram (Fig. 8A).

In the upper boundary of the models a layer with properties of the air was introduced (Table 1) and on the other three sides absorbing boundary conditions were assumed, described by the increase in electrical conductivity (ReflexW Manual 2013).

In the numerical models spatial discretization $\Delta x = \Delta z = 0.01$ m and time discretization $\Delta t = 0.01$ns were applied, according to the criteria of convergence

Table 1. Electromagnetic properties assumed for modelling (Annan 2001, Dortman 1984, Drozdzak and Twardowski 2010)

Tab. 1. Parametry elektromagnetyczne użyte do modelowania (Annan 2001, Dortman 1984, Drozdzak and Twardowski 2010)

Medium	Relative dielectric permittivity ε_r [-]	Electrical conductivity σ [mS/m]	Relative magnetic permittivity μ_r [-]
Gypsum	7	0.001	1
Air	1	0	1
Fresh water	81	0.5	1

and stability of algorithm for the FDTD method. Solving was conducted in time window T = 100 ns adequately to time window assumed during GPR surveys.

The results of numerical modelling, i.e. synthetic radargrams, are shown in Figs. 10B, D and F.

If, at the Siesławice site, a weathered zone did not occur, a radargram should be similar to that shown in Fig. 10B; interpretation of reflections from the roof of the cavity and from the floor (filled with water in analyzed situation) would be possible. The result of modelling confirms that in a homogenous and lossless medium, detection of a void is an easy matter (Fig. 9A and Fig. 10B).

When a weathered zone occurs in the near surface zone a masking effect may mean that interpretation of reflections from voids located at greater depths might be difficult or impossible. In the analyzed situation, when fractures were filled with air, only reflections from the floor (filled with water in analyzed situation) were observed (Fig. 10D); the result of modelling (Fig. 10D) is convergent with the measured radargram (Fig. 8A) where reflections from the floor of the cavity were recorded. The difference in the position of reflections from the floor of the cavity in Fig. 10D (i.e. at z = 6.5 m) and 8A (i.e. at z = 5.5 m) allows the assumption to be made that fractures directly over the roof of the cavity should by partly filled with water; water is a medium that causes a decrease in the velocity of electromagnetic waves and in consequence reflections were recorded at grater depth. The above assumption about only partial saturation of the weathered zone by water was made on the basis of the analysis of the synthetic radargram presented in Fig. 10F. If fractures were fully saturated with water, no reflection below the weathered zone would be recorded in radargrams (Fig. 10F).

The divergence between visualization of weathered zone in measured radargram (Fig. 8A) and synthetic radargrams (Fig. 10D and F) results from the fact that in modelling this zone was defined as homogenous weathered but in fact such assumption might not be correct. Shape of the weathered zone was assumed in the numerical model on the basis of analysis of measured radargram (Fig. 8A). The results of modelling (Fig. 10D and F) deliver information that values of electromagnetic properties and their distribution in the weathered zone influenced stronger to visualization of the weathered zone than shape of this zone.

Synthetic radargrams presented in Fig. 10D and F show an influence of the masking effect on recorded data and this effect results from so-called scattering attenuation. The influence of scattering attenuation on GPR data was discussed, among others, in papers by Annan (2001), Cassidy (2008), Grimm et al. (2006) and Golebiowski (2012). The total value of the attenuation coefficient α is superposition of ohmic attenuation α_{ohmic} (formula 2) and scattering attenuation $\alpha_{scattering}$. Usually in hard rocks ohmic attenuation might be omitted in analysis but scattering attenuation plays an important role due to the presence of natural or/and anthropogenic fractures. In the ground the opposite occurs – the most important is the analysis of ohmic attenuation due to the presence of clay minerals in the ground.

8. Conclusions

As it was presented in the paper georadar and microgravimetric techniques seem to be a useful methods for detection and monitoring of karst processes developing in gypsum formations. Bothe techniques are complementary each other and therefore they deliver

complex information about spatial distribution of fractures and loose zones origin from karst and weathering processes.

Divergences on profiles G-1 and G-2 between modelled microgravimetric curves and measured curves depict to the presence the fractures and loose zones in the region of the cavity. Such interpretation was confirmed by the result of GPR surveys which showed the presence of the near surface weathered zone and its superposition with fractures and loose zones developed over the cavity. Application only microgravimetric technique did not deliver information about presence of weathered zone. From the other hand interpretation of anomalies in radargrams without additional microgravimetric data would be ambiguous, because reflections might be interpreted either as a fractures and loose zones or as a presence in the rock mass different kinds of gypsum formations, etc. Therefore both techniques should be apply together during examination of karst phenomena.

Numerical modelling of the electromagnetic wave field allows it to be stated that in the weathered zone a certain amount of water should occur. Additionally synthetic radargrams allowed to analyses changes in measured data and analyses of possibility of voids detection when weathered zone and karst phenomena are present in the surveys site.

Building a model of cracks and loose zones around the cave requires the removal of the gravity effect of the cave (Golebiowski 2012; Porzucek 2013). The shape of the cave throughout its course is unknown, and the modelled gravity effect was made on the assumption that it has a constant cross section. The calculated effect of gravity is therefore not fully correct, which makes it impossible to create a precise gravimetric model of the rock formation surrounding the cave. Laser scanning inside the cave would allow its exact geometric model to be created, thereby removing the effect of gravity from the Bouguer anomaly. However, implementation of laser scanning would be very difficult to carry out due to the water covering the bottom of the cave.

Geophysical surveys clearly show that both the shape of the cave and the structure of the rock mass surrounding the void are complicated. In this situation only detailed 3D geophysical analysis will deliver full information about the distribution of fractures and loose zones and physical properties of the examined rock mass.

Acknowledgements

The works were conducted during the realization of grants no.: 11.11.140.769 and financed by the AGH University of Science and Technology in Cracow, Poland. Geodetic measurements were carried out by students of the Geodetic Society Dahlta AGH UST.

References

A.S.T.M. 1999. Standard Guide for Selecting Surface Geophysical Methods. Designation D6429. American Society for Testing and Materials, Philadelphia.

Annan A.P. 1999. Practical Processing of GPR Data. Sensor and Software Inc., Canada.

Annan A.P. 2001. Ground Penetrating Radar – Workshop Notes. Sensor and Software Inc., Canada.

Bishop I., Styles P., Emsley S.J., Ferguson N.S. 1997. The Detection of Cavities Using the Microgravity Technique: Case Histories from Mining and Karstic Environments. Geological Society, Engineering Geology Special Publications, No. 12, pp. 153–166.

Blakely R. 1991. Potential Theory in Gravity and Magnetic Applications. Cambridge University Press.

BSI 1999. British Standard Code of Practice for Site Investigations – BS5930. British Standards Institution, London.

Cassidy N.J. 2008. GPR Attenuation and Scattering in a Nature Hydrocarbon Spill: A Modeling Study. Vadose Zone Journal, Vol. 7, No. 1, pp. 140–159.

Chamberlain A.T., Sellers W., Proctor C., Coard R., 2000, Cave Detection in Limestone using Ground

Penetrating Radar. Journal of Archaeol. Science, 27, pp. 957–964.

Dortman N.B. 1984. Physical Properties of Rocks. Niedra Ed., Moscow (in Russian).

Fajklewicz Z. 2007. Applied Gravity, AGH University of Science and Technology Press, Cracow, Poland (in Polish).

Drozdzak R., Twardowski K. 2010. Dielectric Permittivity of Porous Media – Factors Which Influence Its Changes. Drilling Oil and Gas, Cracow, Poland (in Polish).

Flis J. 1954. Gypsum Krast in Nida Gypsum Deposits, Institute of Geography – Polish Academy of Science, Geographic's works no. 1, Warsaw (in Polish).

Golebiowski T. 2004. Introduction to GPR Surveys Interpretation with Procedure of Numerical Modelling. Geological Review, Vol. 52, No. 7, pp. 563–568 (in Polish).

Golebiowski T. 2007. Numerical Modelling of GPR Wave Filed Using of FDTD Method. Geoinformatica Polonica, No. 8, pp. 23–35 (in Polish).

Golebiowski T. 2012. Application of the GPR Technique for Detection and Monitoring of Objects with Stochastical Distribution in the Geological Medium. AGH University of Science and Technology Press, Cracow, Poland (in Polish).

Grimm R.E., Heggy E., Clifford S., Dinwiddie C., McGinnis R., Farrell D. 2006. Absorption and Scattering in Ground Penetrating Radar: Analysis of the Bishop Tuff. Journal of Geophysical Research, Vol. 111, E06S02, 15p.

Jol H.M. 2009. Ground Penetrating Radar: Theory and Applications. Elsevier Science Ed., Oxford, UK.

Karczewski J. 2007. The Principle of GPR Method. AGH University of Science and Technology Press, Cracow, Poland (in Polish).

Kasprzyk A. 1991. Analysis Lithofacial of Sulphate Tracks Builders Southern Margin of the Holy Cross Mountains. Geological Review, no. 4, PIG-PIB Ed., Warsaw, pp. 213–223 (in Polish).

Kasprzyk A. 1993. Lothofacies and Sedimentation of the Badenian (Middle Miocen) Gypsum in the Northern Part of Carphatian Foredeep, Southern Poland. Ann. Soc. Geol. Pol., No. 63, pp. 33–84.

Leucci G., De Giorgi L. 2010. Microgravimetric and Ground-Penetrating Radar Geophysical Methods to Map the Shallow Karstic Cavities Network in a Coastal Area (Marina Di Capilungo, Lecce, Italy). Exploration Geophysics, No. 41, pp. 178–188.

Łaptuś A. 1992. Giant-Scale Cross-bedded Miocene Biocalcarenites in the Northern Margin of the Carpathian Foredeep. Ann. Soc. Geol. Pol., No. 62, pp. 149–172.

Mochales T., Casas A.M., Pueyo E.L., Soriano M.A., Pueyo O., Román M.T., Pocoví A., Ansón D. 2008. Detection of Underground Cavities by Combining Gravity, Magnetic and Ground-Penetrating Radar Surveys: a Case Study from the Zaragoza Area, NE Spain, Environmental Geology, 53, pp. 1067–1077.

Nowak W.A. 1986a. Nida Gypsum Deposits Physiographic Regionalization. SODF Ed., No. 14, pp. 23–33 (in Polish).

Nowak W.A. 1986b. Krast in Nida Gypsum Deposits. SODF Ed., No. 14, pp. 87–117 (in Polish).

Porzucek S. 2013. Loosenings and Cracks Detection in Rock Mass Located Above Anthropogenic Voids Using the Microgravity Method, AGH University of Science and Technology Press, Cracow, Poland (in Polish).

ReflexW Manual 2013. SandmeierGeo. Karlsruhe, Germany.

Rutkowski J. 1986. Geological Structure of Nida Gypsum Deposits. SODF Ed., No. 14, pp. 35–61 (in Polish).

Rybakov M., Goldshmidt V., Fleischer L., Rotstein Y. 2001. Cave Detection and 4-D Monitoring: a Microgravity Case History Near the Dead Sea. The

Leading Edge (Society of Exploration Geophysicists), 20(8), pp. 896–900.

Sharma P.V. 2004. Environmental and Engineering Geophysics. Cambridge University Press, Cambridge.

Telford W.M., Geldart L. P., Sheriff R. E. 2004. Applied Geophysics, 2nd ed., Cambridge University Press, Cambridge.

Turchinov I.I. 1997. Lithological conditions of the development of karst processes in the Carpathian Foredeep Badenian gypsum. Przegląd Geologiczny, 45, pp. 803–806 (in Polish).

Walaszczyk I. 1992. Turonian Through Santonian Deposit of the Central Polish Uplands, Their Facies Development, Inoceramid Paleontologie and Stratigraphy. Acta Geol. Pol., 42, 1–2, pp. 1–122.

Wołoszyn B.W. 1990. Cave in Ponidzie – Team of Landscape Parks. SODF Ed., No. 18, pp. 275–341 (in Polish).

Yee K. 1966. Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media. Transactions on Antennas and Propagation, No. 14, IEEE Ed., pp. 302–307.