Mining Science, vol. 20, 2013, 5−17

www.gornictwoigeologia.pwr.wroc.pl

Mining Science
(previously Prace Naukowe Instytutu Górnictwa Politechniki Wrocławskiej. Górnictwo i Geologia)

ISSN 2300-9586
previously 0370-0798

Received July 16, 2013; reviewed; accepted September 27, 2013

multivariate statistical analysis, self-organising maps, mining deformation, GIS

Jan BLACHOWSKI[1], Paweł STEFANIAK[2]

APPLICATION OF THE SPATIAL DATA MINING MODULE IN ANALYSIS OF MINING GROUND DEFORMATION FACTORS

Spatial data mining methods for example those based on artificial neural networks (ANN) allow extraction of information from databases and detection of otherwise hidden patterns occurring in these data and in consequence acquiring new knowledge on the analysed phenomena or processes. One of these techniques is the multivariate statistical analysis, which facilitates identification of patterns otherwise difficult to observe.

In the paper an attempt of applying self-organising maps (SOM) to explore and analyse spatial data related to studies of ground subsidence associated with underground mining has been described. The study has been carried out on a selected part of a former underground coal mining area in SW Poland with the aim to analyse the influence of particular ground deformation factors on the observed subsidence and the relationships between these factors. The research concerned the uppermost coal panels and the following factors: mining system, time of mining activity and inclination, thickness and depth below the ground of the exploited coal panels.

It has been found that the exploratory spatial data analysis can be used to identify relationships in multidimensional data related to mining induced ground subsidence.

The proposed approach may be found useful in identification of areas threatened by mining related subsidence and in creating scenarios of developing deformation zones and therefore aid spatial development of mining grounds.

[1] Wroclaw University of Technology, Institute of Mining Engineering, Na Grobli 15, 50-421 Wrocław, Poland.

[2] Wroclaw University of Technology, The Faculty of Geoengineering, Mining and Geology, Na Grobli 15, 50-421 Wrocław, Poland.

doi: 10.5277/gig132001

1. INTRODUCTION

Mining grounds subsidence is a time-dependent deformation of the surface caused by readjustment of the overburden above the space that have been created by underground mining. Ground subsidence in an area of abandoned mining is due to past underground mining activity. The subsidence is a process that is influenced by multiple factors associated with underground mining. The studies and description of the deformations that occur on the surface require analysis of the relationships between these factors and the registered ground movements. The major factors influencing ground deformations may include: geological and tectonic conditions and the associated geometry of the deposit (e.g. height of the production levels) and mining systems used to extract the mineral, location of the deposit with respect to the ground surface (e.g. depth below the surface and inclination of the excavated levels), and rate of mining. The relationships between these factors are complicated, therefore their influences and their combinations need to be systematically analysed.

The scientific understanding of complex spatial problems associated with deformation studies often depends on the: (a) discovery, (b) interpretation, and (c) presentation of multivariate patterns, e.g. detection of unknown multivariate spatial relationships between the incidence factors of rock mass activity. Nowadays information about these factors is usually stored in geographical databases and the analyses are facilitated with geographical information systems (GIS). The most significant, in the scope of this paper, attempts of analysing mining deformations with the aid of geoinformation systems described in literature include work by (Oh et al, 2010) that describes methodology of developing ground subsidence hazard maps near abandoned coal mines using frequency ratio and sensitivity analysis to produce ground subsidence hazard index with the aid of GIS. The authors have analysed the significance of the following factors on the observed ground subsidence: depth of mining, terrain slope, land use, distance from drift, depth of groundwater, geology, rock mass rating, permeability and lineament. Oh and Lee (2010) assessed the spatial ground subsidence hazard by using a GIS with models based on frequency ratio (FR), logistic regression (LR) and an artificial neural network (ANN). Djamaluddin et al. (2011) have used GIS to facilitate subsidence prediction with theoretical method using the Knothe function and apply fuzzy logic to identify damage classification zones. Whereas Zahiri et al. (2006) applied the weights-of-evidence (WOE) method to derive rock fall potential associated with mining induced subsidence. Choi et al. (2010) constructed subsidence susceptibility maps based on fuzzy relations for an abandoned underground coal mine. All but the first one of this research provides the predicted ground subsidence area as hazard maps and provide little information about the assessment of the effect of ground subsidence-related factors.

In this paper, a multivariate spatial analysis based on the Kohonen network (Kohonen, 2001), also known as self-organizing maps (SOM) has been performed to determine the influence of factors associated with underground mining affecting surface subsidence. SOM is a form of artificial neural network (ANN) that is trained using unsupervised learning to produce a two-dimensional discretized representation of the input space of the training samples. SOM are different from other artificial neural networks in the sense that they use a neighbourhood function to preserve the topological properties of the input space.

The purpose of this study was to analyse the relationships between the ground subsidence area and geological and mining factors, calculate weights for these factors and provide a method to predict areas of ground subsidence after the end of mining using GIS and SOM software packages. The ESRI ArcGIS v.10 GIS software has been used for spatial data handling, which is also the environment of the geoinformation system of the former mines. There are various, both public domain and commercial software packages available for data analysis and visualization with SOM. For example the *SOM_PAK* developed by the SOM Programming Team of the Helsinki University of Technology (Kohonen et al. 1996). In this study the *VIS-STAMP: A Visualization System for Space-Time and Multivariate Patterns* package has been used. It couples computational, visual, and cartographic methods for exploring and understanding spatio-temporal and multivariate data and has been created at the Spatial Data Mining and Visual Analytics Lab of Department of Geography University of South Carolina (Guo et al., 2006).

The work has been carried out with the aid of the geoinformation system used for facilitating ground deformation studies that comprises of a geographical database and modules for data acquisition and management, spatial data mining and modelling and visualisations (Blachowski 2008, Blachowski and Stefaniak 2012).

2. STUDY AREA

The methodology of multivariate spatial analysis for mining deformation studies has been applied on a selected part of the former mining grounds in and around the city of Wałbrzych (SW Poland). This area was subjected to underground exploitation of hard coal that lasted for several hundred years and ended in the late 90-ties of the 20th Century. The effects of the coal mining industry are still visible in the local landscape in the form of dumps, settlement ponds and subsidence basins. The results of geodetic measurements carried out after the end of mining indicate presence of secondary ground deformations (Blachowski et al, 2010). The area selected for this study, the Barbara field of the Victoria coal mine, one of the four mines in the area that had been operating there after the World War II, has been shown in Fig. 1. The area lies between the X = 301700 m and 304700 m and Y = 322 000 m and 325 000 m coordinates of the PUWG 1992 coordinate system. The analysed part of the Barbara field

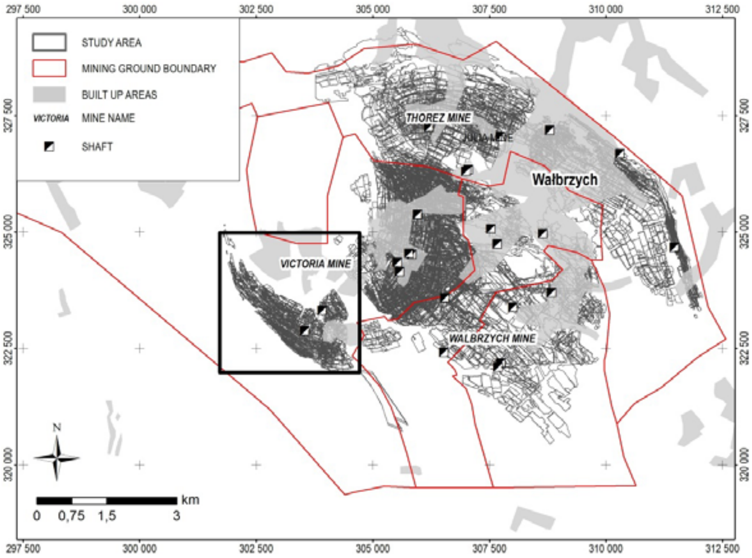

Fig.1. Location of the study area – the Barbara coal field – in the former Walbrzych Coal Basin

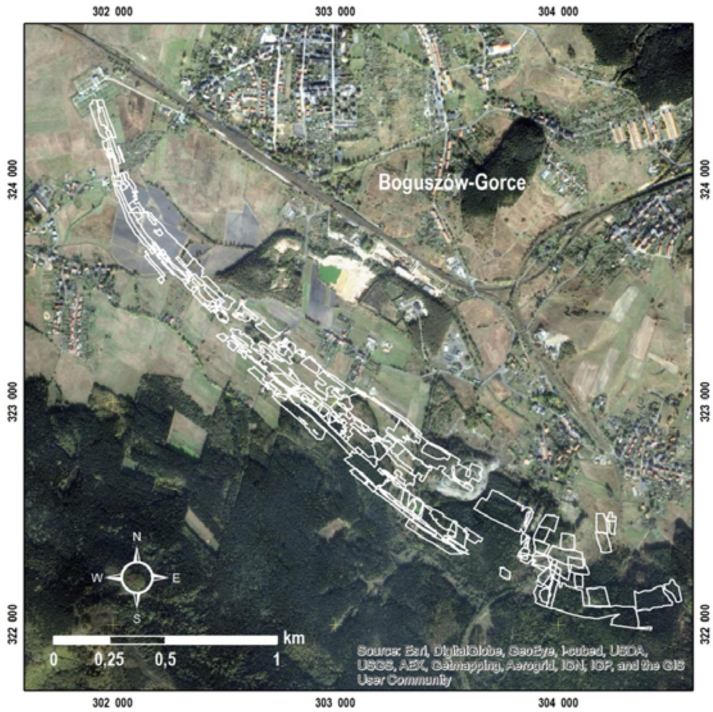

Fig. 2. Aerial photo of the study area showing present-day land use
and outline of the analysed coal panels

has 16 coal levels that had been mined throughout the 19th and the 20th Centuries. Two main coal mining methods were used, longwall and caving and longwall with various forms of fill. These coal levels are associated with the *Zaclerskie* coal bearing formations, which are inclined at 30 to 60 degrees to the horizontal. The thickness of the coal levels varies from 0.7 to 2.2 m, sometimes the coal deposits are thinning or disappear completely.

3. MATERIALS AND METHODS

3.1. DATA AND PREPARATION

The initial processing of data in a GIS concerned extraction of vector type spatial data from the geoinfomation system database and their conversion to a grid type data for multivariate spatial data analysis. The study area, 3 x 3 km (Fig. 2), has been divided into 50x50 m cells (Fig. 3). These cells have been then assigned the values of particular factors based on their spatial relationship with vector data representing geometry of the underground workings and the observed ground subsidence. The vector data, stored in the geoinformation system database have been compiled based on 1:5000 scale scanned mining maps. The following subsidence factor maps in the grid format have been prepared:

- coal level thickness,
- time of mining,
- mining system used,
- inclination of coal production level,
- depth of coal level below the surface,
- ground subsidence areas.

The first three maps have been produced based on attribute data associated with the polygons representing underground workings using the GIS data conversion functions. The resulting maps are in raster format with 50 m cell resolution. To obtain the inclination, triangulated irregular network (TIN) model representing surface of particular coal level has been produced from the vector data and the inclination has been calculated using the GIS slope analytical function. The depth of coal level below the ground surface has been calculated by subtraction of the coal production level surface raster from the ground surface raster (DEM) using the map algebra functionality. The map algebra treats spatial datasets as spatial variables, to which algebraic operations can be applied (Tomlin, 2006). The ground subsidence map has been produced from vector polygon data obtained from (Kowalski, 2000). The individual factor maps have been combined into a single dataset representing parameters for a given coal production level. For the initial analysis the top coal production levels have been selected. For analysis of multiple coal levels the values of individual factors have been averaged (e.g. inclination) or summed up (e.g. thickness). The fig. 3 shows factor maps with parameter values for the coal level 306/307.

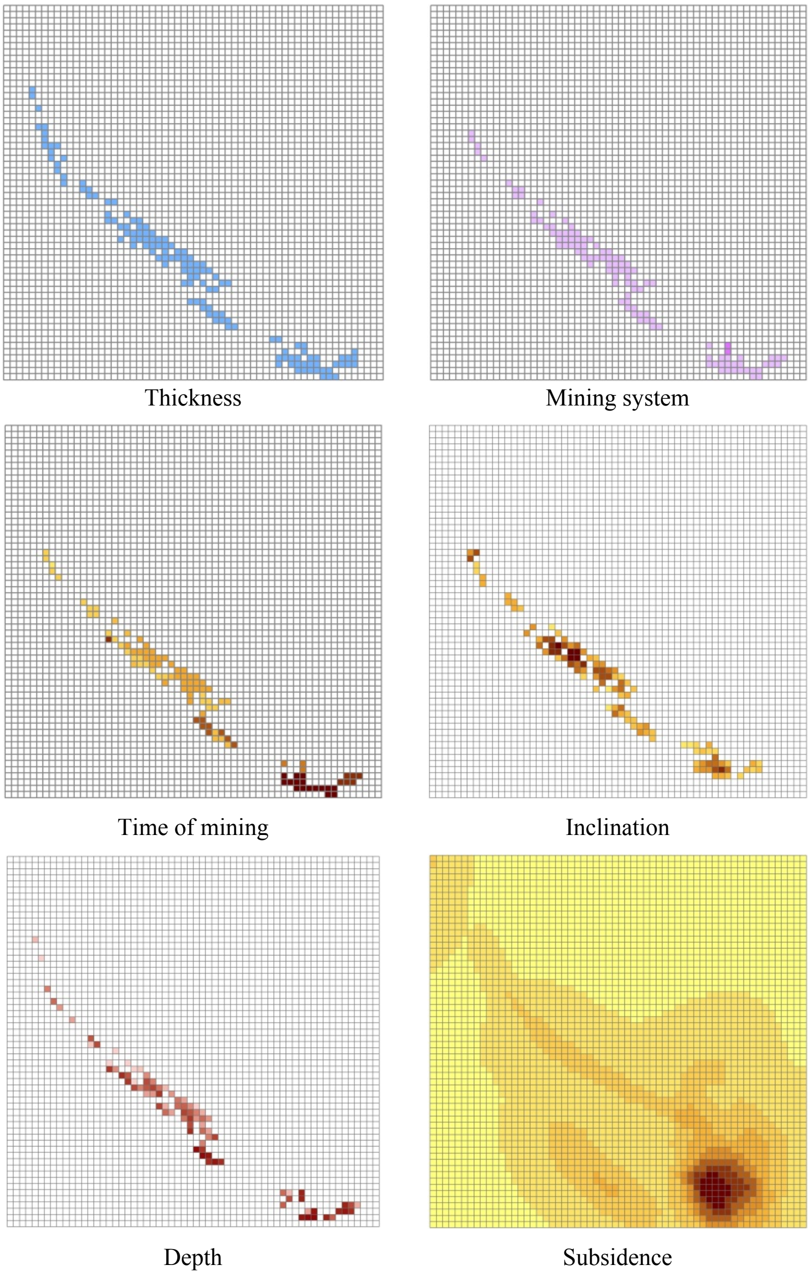

Fig. 3. Factor maps for level 306/307 (darker colours indicate higher values)

The data extracted from the GIS database as input for multivariate statistical analysis has the form of a cube that is defined by three parameters: a) the spatial reference (i.e. the study area represented as GRID), the time (i.e. mining activity) and a set of variables (i.e. the deformation factors) represented by floating point values (fig. 4).

Each element of the data cube is a particular spatial entity we call for the purpose of this study a *GeoCell* associated with a particular input GRID cell, particular period in a time series and a particular variable (e.g. thickness, inclination, depth of coal panel or mining system used). A horizontal layer of the data cube is assigned to a particular spatial entity a time-attribute slice (Fig. 5).

The input dataset consists of 205 geo-cells, 6 variables (factors) and up to 10 time periods in the 1800-1996 time series for each of the coal production levels analysed. Therefore there are 205 time –attribute slices and 2050 (205x10) multivariate profiles for a given time period.

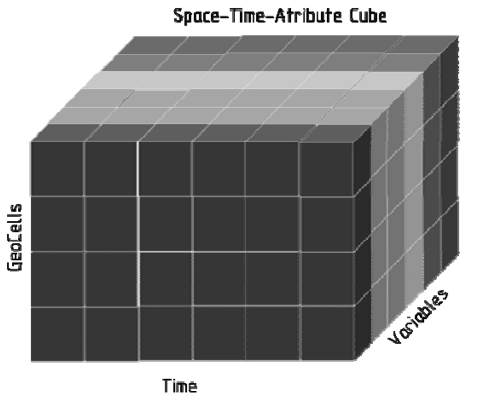

Fig. 4. Visualisation of the data cube for the multivariate statistical analysis

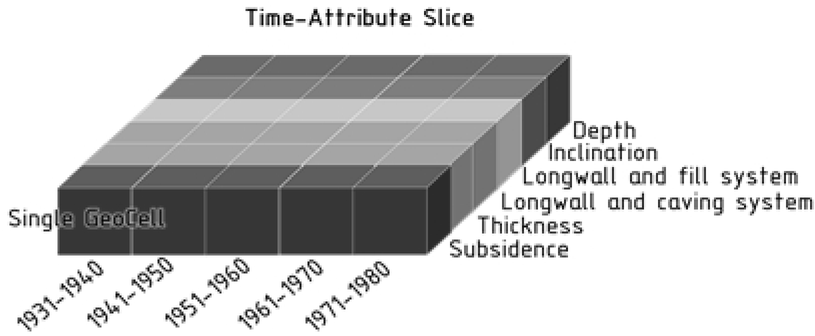

Fig. 5. Visualisation of the time– attribute slice in the data cube

3.2. METHODOLOGY

The SOM is constructed from nodes, which are organized as a regular 2-dimensional map made up of squares or hexagons. These represent topology and neighbourhood of neurons forming clusters of data in multidimensional space A single neuron is represented by codebook (weight) vector $m=[m_1, m_2, m_3,\ldots m_d]$ from d-dimensional space where d is the number of variables of the input data vectors x. The neurons are connected with nearby neurons based on their neighbourhood relation, which determine the topology and structure of the map. The structure of the SOM consists of regular M units, the inside of which are the reference vectors. These represent individual input vectors and indicate their individual codebook vectors (and conversely).

The SOM is trained iteratively. In a single training step, one sample data vector x from the input dataset is chosen randomly and presented by the input layer to network. Then, the distances between this data vector and all the codebook vectors are calculated using Euclidean geometry. This is the measure of similarity of the input data vectors. If a codebook vector is the closest to the presented input data vector x than its neuron is the winner called the Best-Matching Unit (BMU). The BMU has a precedence to adapt its weights. The SOM algorithm can be divided into two steps. In the first step of training the network called the *WTM – Winner Takes Most*, the radius of the neighbourhood search is large, thus the BMU and its neighbours in the map change their codebook vectors by updating their weights. The weights are updated following the expression (1) (Kohonen, 2001),

$$w_i(k+1) = w_i(k) + \eta(k)\cdot h_{ci}(k)\cdot\left[x(k) - w_i(k)\right] \tag{1}$$

Where: w_i – is the weight of the i-neuron,
K – is the number of iteration,
$\eta(k)$ – is the learning parameter,
c – is the number of the BMU, in the k-iteration,
$h_{ci}(k)$ – is the parameter assigning the membership of the i-neuron to the neighbourhood of c,
x – is the input vector.

In this step the learning parameter is too large. The second step is called the *WTA – Winner Takes All*. Here, the radius of the neighbourhood converges to zero, thus only the BMU's update their weights. The change of the codebook vector is performed by using the following expression (Kohonen, 2001):

$$w_{c`}(k+1) = w_c(k) + \eta(k)\left[x - w_c(k)\right] \tag{2}$$

The SOM training process has been shown in fig. 6. In the beginning each neuron recognises its class and the multidimensional space is divided into equal regions e.g.

the Voronoi polygons. The neighbourhood of reference vectors (left) represents, topologically, the neighbourhood of the data vectors (right).

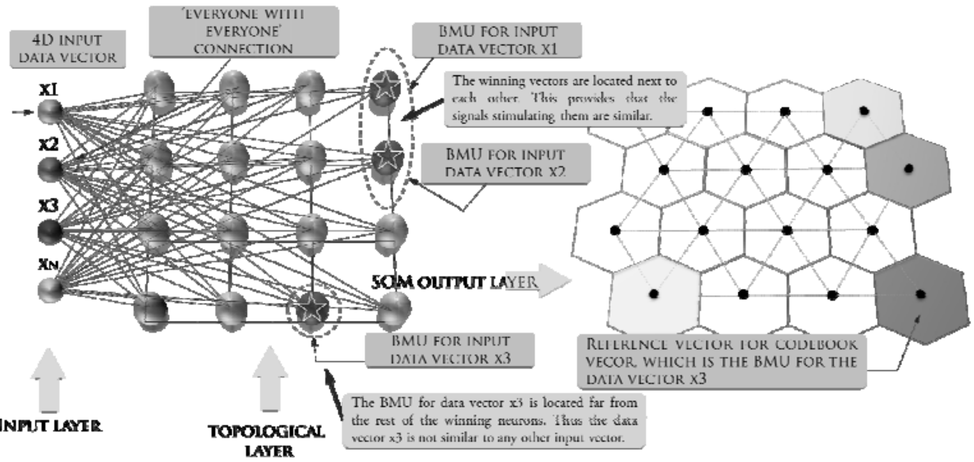

Fig. 6. The process of SOM learning (after Gramacki and Gramacki, 2008)

The results of data analysis and clustering can be visualized in the form of *the U-Matrix*, which is the typical SOM projection or the *Parallel Coordinate Plot – PCP*. The latter is a method for presenting a set of multivariate data points in two-dimensional form. Each data point is represented by a horizontal segment line connecting vertical axes corresponding to subsequent variables (Szustalewicz, 2002).

The *U-Matrix* presents SOM results using two different types of hexagons: (a) node hexagons and (b) distance hexagons. Each of the node hexagons is shaded to represent distance (dissimilarity) between all the neighbouring codebook vectors and contains a point, which has assigned a specific colour. The points are scaled to show the number of input data vectors in the node but the color indicates specific patterns occurring in the data. The shading of the distance hexagons represents the multivariate distance between two neighbouring codebook vectors - lighter colours represent closely spaced node codebook vectors and darker colours indicate more distant ones.

4. RESULTS

The above described analytical approach, using the self-organizing maps methodology, has been used to analyse and visualise the relationships and influence of factors associated with mining ground deformations. The analysis involved quantitative data related to the observed ground subsidence and mining parameters such as: inclination, depth, thickness and mining method for two coal production levels in 10 year periods between the 1880 and 1980.

To examine the patterns in the analysed data *U-Matrix* and *PCP* visualisations have been used. Two distinct patterns have been found. The first one concerns a relationship between the following factors: inclination and thickness of coal panels and the longwall and caving mining system and the observed subsidence. These are related to the greatest values of subsidence, identified in the SE part of the analysed mining area and the 1940-1960 period. This has been shown in fig. 7 with the selected SOM node.

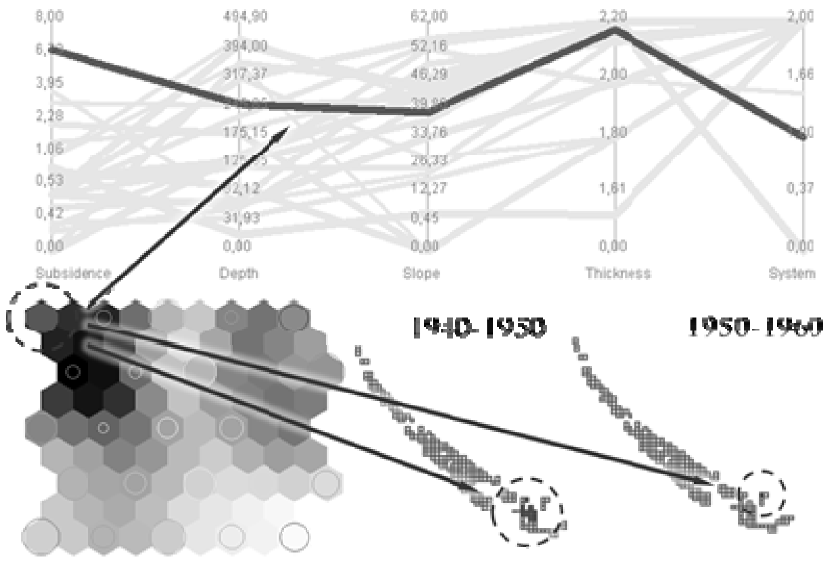

Fig. 7. Visualisation of the relationship between the observed subsidence and the analysed subsidence factors

The second pattern concerns the analysis of the longwall and fill mining system used in shallow coal panels. For this purpose a selection in the *U-Matrix* of the SOM nodes representing this mining system and above average values of inclination and thickness variables has been made. The results shown in fig. 8 indicate that there is association of this mining system and low values of subsidence. These are most clear in the 1900–1910 period where it can be observed for the NW part of the analysed coal field.

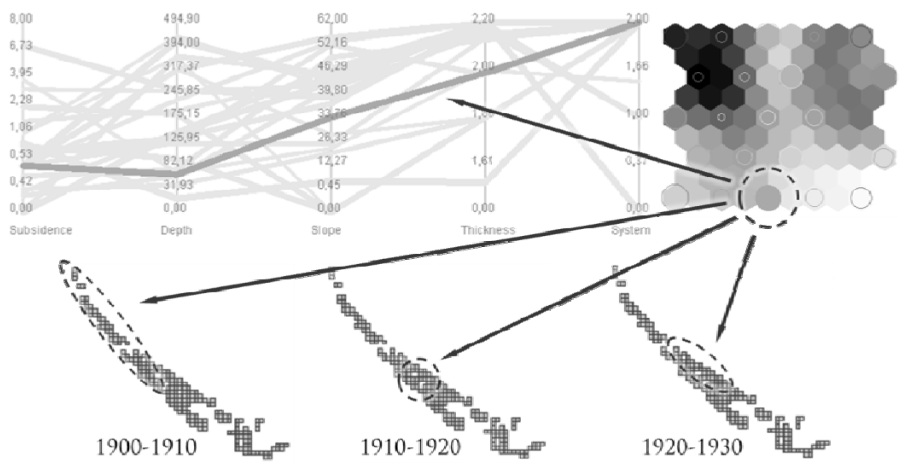

Fig. 8. Visualisation of the analysis for the longwall and fill mining system used in shallow coal panels and the ground subsidence relationship

5. CONCLUSIONS

The paper describes the results of preliminary studies concerning application of the SOM methodology to analyse and identify relationships between mining ground deformation factors and the observed ground subsidence. To test the SOM applicability for multivariate statistical analysis of data related to mining subsidence and their visualisations input data related to one of the coal fields *Barbara* of the former Wałbrzych Coal Basin have been prepared in the geoinformation system (Blachowski 2008, Blachowski and Stefaniak, 2012). The following factors have been used as input variables for analysis: inclination, thickness and depth of the exploited coal panels, as well as the mining system used and time of mining. The initial study concerned the top coal panels. It has been proved on the example of the mining deformation factors (e.g. mining system) and ground subsidence analysis that exploratory spatial analysis can be used to discover relationships in multidimensional data. Such information can be used in studies of ground deformation processes and identification of zones threatened by mining subsidence. It must however be remembered that these initial results need to be subjected to further tests. This will be done including the remaining coal panels mined in the area.

ACKNOWLEDGMENTS

This work has been done with the research project S20041 *"Development of the secondary deformations model on mining grounds"* (J. Blachowski) and the European Union as part of the European Social Fund (P. Stefaniak).

LITERATURE

BLACHOWSKI J., 2008. *System informacji geograficznej wałbrzyskich kopalń węgla kamiennego podstawą zwiększenia efektywności i wiarygodności badań deformacji powierzchni terenów pogórniczych*, Pr. Nauk. Inst. Gór. PWroc., Stud. Mater., nr 123, Górnictwo i geologia X, nr 34, 17–27.

BLACHOWSKI J., STEFANIAK P., 2012. *Aktualizacja systemu geoinformacyjnego dawnych Wałbrzyskich Kopalń Węgla Kamiennego*, Pr. Nauk. Inst. Gór. PWroc., Stud. Mater., nr 135, Górnictwo i geologia XVIII, nr 42, 5–21.

BLACHOWSKI J., MILCZAREK W., CACOŃ S., 2010. *Project of a rock mass surface deformation monitoring system in the Walbrzych coal basin*, Acta Geodynamica et Geomaterialia, Vol. 7, No 3, 349–354.

CHOI J.K., KIM K.D., LEE S., WON J.S., 2010. *Application of a fuzzy operator to susceptibility estimations of coal mine subsidence in Taebaek City, Korea*, Environ Earth Sci 59(5):1009–1022.

DJAMALUDDIN I., MITANI Y., ESAKI T., 2011. *Evaluation of ground movement and damage to structured from Chinese coal mining using a new GIS coupling model*, International Journal of Rock Mechanics and Mining Sciences, Vol. 48, 3, 380–393.

GRAMACKI J., GRAMACKI A., 2008. *Wybrane metody redukcji wymiarowości danych oraz ich wizualizacji*, XIV Konferencja PLOUG, Szczyrk.

GUO D., CHEN J., MACEACHREN A. M., LIAO K., 2006. *A Visualization System for Spatio-Temporal and Multivariate Patterns (VIS- STAMP)*, IEEE Transactions on Visualization and Computer Graphics, 12(6), 1461–1474.

KOHONEN T., HYNNINEN J., KANGAS J., LAAKSONEN J., 1996. *SOM_PAK: The Self-Organizing Map Program Package*, Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland

KOHONEN T., 2001. *Self-Organizing Maps*. 3rd Edition. Springer.

KOWALSKI A. (red.), 2000. *Eksploatacja górnicza a ochrona powierzchni. Doświadczenia z wałbrzyskich kopalń*, Główny Instytut Górnictwa, Katowice.

OH H.J., LEE S., 2010. *Assessment of ground subsidence using GIS and the weights-of-evidence model*, Engineering Geology, Vol. 115, 1–2, 36–48.

OH H.J., AHN S.Ch., CHOI J.K., LEE S., 2011. *Sensitivity analysis for the GIS-based mapping of the ground subsidence hazard near abandoned underground coal mines*, Environ Earth Sci 64:347–358.

STEFANIAK P., ZIMROZ R., 2013. *Multivariate diagnostic data analysis from gearboxes of spatial distributed conveying system*, Interdisciplinary Topics in Mining and Geology, edited by J. Drzymala and W. Ciezkowski, Oficyna Wydawnicza Politechniki Wrocławskiej.

SZUSTALEWICZ A., 2002. *Visualisation of multivariate data using parallel coordinate plots and Kohonen's SOM's. Which is better?* Advanced Computer Systems, The Springer International Series in Engineering and Computer Science, Vol. 664, pp. 89-98.

TOMLIN C. D., 2006. *Cartographic Modeling*, In: Encyclopedia of Geographic Information Science. Sage.

ZAHIRI H., PALAMARA D.R., FLENTJE P., BRASSINGTON G.M., BAAFI E., 2006. *A GIS-based weights-of-evidence model for mapping cliff instabilities associated with mine subsidence*. Environ Geol. 51(3), 377–386.

ZASTOSOWANIE MODUŁU EKSPLORACJI DANYCH PRZESTRZENNYCH W ANALIZACH CZYNNIKÓW DEFORMACJI TERENÓW GÓRNICZYCH

Metody eksploracji danych przestrzennych na przykład te oparte na sztucznych sieciach neuronowych (SSN) pozwalają na ekstrakcję informacji z baz danych i wykrywanie ukrytych relacji występujących w tych danych, a w konsekwencji pozyskiwanie nowej wiedzy o analizowanych zjawiskach i procesach. Jedną z grup technik eksploracji danych przestrzennych jest statystyczna analiza wielowymiarowa (ang. *multivariate statistical analysis*), która umożliwia identyfikację wzorców inaczej trudnych do wykrycia. W pracy przedstawiono próbę zastosowania metodyki samoorganizujacyh się map (SOM) w eksploracji i analizie danych przestrzennych na potrzeby wspomagania badań deformacji powierzchni spowodowanych podziemną działalnością górniczą. Badania przeprowadzono na wybranym fragmencie dawnego zagłębia węgla kamiennego w Polsce w celu analiz wpływu czynników deformacji górotworu na obserwowane osiadania powierzchni i związków między tymi czynnikami. Dotyczyły one dwóch górnych pokładów węgla i następujących czynników: system eksploatacji, okres eksploatacji, nachylenie, miąższość i głębokość eksploatowanych pokładów poniżej powierzchni terenu.

W wyniku przeprowadzonych badań stwierdzono przydatność metody SOM do identyfikacji związków w danych wielowymiarowych dotyczących deformacji terenów górniczych

Proponowane podejście może także znaleźć zastosowanie w identyfikacji obszarów zagrożonych osiadaniami oraz w budowaniu scenariuszy rozwoju stref deformacji, a przez to wspomaganie planowania zagospodarowania przestrzennego takich obszarów.