www.czasopisma.pan.pl PAN www.journals.pan.pl

DE GRUYTER OPEN

Arch. Min. Sci., Vol. 61 (2016), No 3, p. 523–535

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

DOI 10.1515/amsc-2016-0038

DARIUSZ JAMRÓZ*, TOMASZ NIEDOBA**, AGNIESZKA SUROWIAK**,
TADEUSZ TUMIDAJSKI**

THE USE OF THE VISUALISATION OF MULTIDIMENSIONAL DATA USING PCA TO EVALUATE POSSIBILITIES OF THE DIVISION OF COAL SAMPLES SPACE DUE TO THEIR SUITABILITY FOR FLUIDISED GASIFICATION

ZASTOSOWANIE WIZUALIZACJI WIELOWYMIAROWYCH DANYCH ZA POMOCĄ PCA DO OCENY MOŻLIWOŚCI PODZIAŁU PRÓBEK WĘGLA ZE WZGLĘDU NA ICH PRZYDATNOŚĆ DO ZGAZOWANIA

Methods serving to visualise multidimensional data through the transformation of multidimensional space into two-dimensional space, enable to present the multidimensional data on the computer screen. Thanks to this, qualitative analysis of this data can be performed in the most natural way for humans, through the sense of sight. An example of such a method of multidimensional data visualisation is PCA (*principal component analysis*) method. This method was used in this work to present and analyse a set of seven-dimensional data (selected seven properties) describing coal samples obtained from Janina and Wieczorek coal mines. Coal from these mines was previously subjected to separation by means of a laboratory ring jig, consisting of ten rings. With 5 layers of both types of coal (with 2 rings each) were obtained in this way. It was decided to check if the method of multidimensional data visualisation enables to divide the space of such divided samples into areas with different suitability for the fluidised gasification process. To that end, the card of technological suitability of coal was used (Sobolewski et al., 2012; 2013), in which key, relevant and additional parameters, having effect on the gasification process, were described. As a result of analyses, it was stated that effective determination of coal samples suitability for the on-surface gasification process in a fluidised reactor is possible. The PCA method enables the visualisation of the optimal subspace containing the set requirements concerning the properties of coals intended for this process.

Keywords: Principal Component Analysis, multidimensional visualisation, coal gasification, jigging

* AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, FACULTY OF ELECTRICAL ENGINEERING, AUTOMATICS, COMPUTER SCIENCE AND BIOMEDICAL ENGINEERING, DEPARTMENT OF APPLIED COMPUTER SCIENCE, AL. MICKIEWICZA 30, 30-059 KRAKÓW, POLAND, E-MAIL: jamroz@agh.edu.pl

** AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, FACULTY OF MINING AND GEOENGINEERING, DEPARTMENT OF ENVIRONMENTAL ENGINEERING AND MINERAL PROCESSING, AL. MICKIEWICZA 30, 30-059 KRAKÓW, POLAND, E-MAIL: tniedoba@agh.edu.pl, asur@agh.edu.pl, tadeusz.tumidajski@agh.edu.pl

Proces zgazowania węgla jest jedną z technologii, które zyskują coraz szerszą uwagę wśród technologów zajmujących się jego przeróbką i utylizacją. Ze względu na typ zgazowania wyróżnia się dwa główne sposoby: zgazowanie naziemne i podziemne. Każdy z tych typów można jednak przeprowadzić za pomocą różnych technologii. W przypadku zgazowania naziemnego, jedną z takich technologii jest zgazowanie w reaktorze fluidalnym. Do tego typu zgazowania zostały opracowane wytyczne w ramach projektu NCBiR nr 23.23.100.8498/R34 pt. „Opracowanie technologii zgazowania węgla dla wysokoefektywnej produkcji paliw i energii" w ramach strategicznego programu badań naukowych i prac rozwojowych pt. „Zaawansowane technologie pozyskiwania energii" (Marciniak-Kowalska, 2011-12; Sobolewski et al., 2012; 2013; Strugała et al., 2011; 2012). Autorzy wybrali główne z tych wytycznych, dotyczących zalecanych poziomów określonych cech węgla. W celu zbadania węgla pod kątem ich przydatności do zgazowania pobrano próbki dwóch węgli: pochodzących z Zakładu Górniczego Janina oraz z Kopalni Węgla Kamiennego Wieczorek. Każdy z tych węgli został poddany procesowi wzbogacania w laboratoryjnej osadzarce pierścieniowej (10 pierścieni, węgiel w klasach wydzielonych z przedziału 0-18 mm). Po zakończeniu procesu rozdziału materiał podzielono na 5 warstw (po 2 pierścienie) i każdy z nich rozsiano na sitach na 10 klas ziarnowych, ustalając wychody warstw i klas. Następnie, tak otrzymane produkty – klasy ziarnowe, po wydzieleniu analitycznych próbek, poddano chemicznej analizie elementarnej i technicznej węgla, w celu scharakteryzowania właściwości wpływających na procesy zgazowania. Łącznie z obu kopalń uzyskano 99 próbek (50 z kopalni Janina oraz 49 z kopalni Wieczorek – w jednej z warstw nie uzyskano klasy 16-18 mm) charakteryzowanych przez następujące parametry: zawartość siarki całkowitej, zawartość wodoru, zawartość azotu, zawartość chloru, zawartość węgla całkowitego, ciepło spalania oraz zawartość popiołu. Przykładowe dane dla jednej z otrzymanych warstw przedstawiono w tabeli 1. Dodatkowo wykorzystano kartę przydatności technologicznej węgla (Sobolewski et al., 2012; 2013), w której opisano parametry kluczowe, istotne oraz dodatkowe, mające wpływ na proces zgazowania. Na jej podstawie oznaczono próbki węgla, które w sposób efektywny poddają się procesowi zgazowania.

W celu wizualizacji danych zastosowano jedną z nowoczesnych metod wielowymiarowej statystycznej analizy czynnikowej – metodę PCA (ang. *Principal Component Analysis*). W metodzie tej dokonuje się rzutu prostopadłego wielowymiarowych danych na płaszczyznę reprezentowaną przez specjalnie wybrane wektory V_1,V_2. Są to wektory własne, odpowiadające dwóm największym (co do modułu) wartościom własnym macierzy kowariancji zbioru obserwacji. Opisany dobór wektorów V_1,V_2 pozwala uzyskać obraz na płaszczyźnie prezentujący najwięcej zmienności danych. Algorytm i zasady tej metody zostały szczegółowo zaprezentowane w podrozdziale 3 artykułu.

Za pomocą metody PCA dokonano trzech typów analiz. Pierwszy obraz miał na celu rozpoznanie, czy możliwa jest identyfikacja pochodzenia węgla, czyli rozdział węgla pochodzącego z ZG Janina od węgla z KWK Wieczorek. Odpowiedź była twierdząca. Na tak przygotowane dane narzucono następnie warunki wynikające z nałożenia wymogów określonych w karcie przydatności technologicznej węgla. Okazało się, że przy wzięciu pod uwagę wszystkich warunków jedynie 17 próbek z ZG Janina i zaledwie jedna z KWK Wieczorek spełnia wszystkie kryteria, co przedstawiono na rysunku 2. Stwierdzono, że dzieje się tak głównie z powodu zawartości chloru, która wykracza poza nałożone limity. Cecha ta nie wpływa jednak w kluczowy sposób na sam proces zgazowania a istotna jest ze względu na aspekt ochrony środowiska. Dlatego dokonano podobnej analizy, ale przy odrzuceniu warunku dotyczącego tej cechy. Po odrzuceniu wymogów dotyczących zawartości chloru okazało się, że 37 próbek z ZG Janina oraz 41 próbek z KWK Wieczorek spełnia pozostałe zalecenia odnośnie naziemnego zgazowania w reaktorze fluidalnym. Jest to potwierdzenie wcześniejszych obserwacji autorów w tym zakresie. W obu przypadkach wizualizacja wielowymiarowa przy użyciu PCA pozwoliła stwierdzić, że obrazy punktów reprezentujących próbki węgla bardziej podatnego na zgazowanie oraz mniej przydatnego do zgazowania zajmują osobne podobszary przestrzeni oraz gromadzą się w skupiskach, które można łatwo od siebie odseparować. Stwierdzono więc, że metoda PCA pozwala podzielić przestrzeń próbek na obszary o różnej przydatności do procesu zgazowania fluidalnego zarówno gdy przyjęto ograniczenie dotyczące zawartości chloru jak i przy jego pominięciu. Zastosowanie metody PCA w celu identyfikacji przydatności próbek węgla do zgazowania jest nowatorskie i nie było wcześniej stosowane. Istnieje możliwość zastosowania również innych metod w tym zakresie. Należy jednak podkreślić, że niewątpliwą zaletą metody PCA jest fakt, że w trakcie wizualizacji nie ma konieczności doboru żadnych parametrów w przeciwieństwie do wielu innych metod wizualizacji wielowymiarowych danych.

Słowa kluczowe: analiza PCA, wizualizacja wielowymiarowa, zgazowanie węgla, wzbogacanie w osadzarkach

1. Introduction

In the article, coal samples coming from two coal mines – Janina Mining Plant and Wieczorek Coal Mine were analysed. The above mentioned samples were taken in order to evaluate their suitability for on-surface gasification in fluidised bed. The properties of coals directed to gasification must comply with the limits (Blaschke, 2009; Borowiecki et al., 2008; Chmielniak & Tomaszewicz, 2012; Kosminski et al., 2006; Lee et al., 2007; Marciniak-Kowalska, 2012-13; Strugała et al., 2011; Strugała & Czerski, 2012; Surowiak, 2013a,b; 2014a,b), wherein it must be noted that they are linked with each other. The evaluation of coal suitability for gasification should be therefore conducted multidimensionally with the use of multidimensional distributions of properties and their statistics (Ahmed & Drzymała, 2005; Brożek & Surowiak, 2005, 2007; 2010; Drzymała, 2009; Jamróz, 2009; 2014a-c; Jamróz & Niedoba, 2014; Niedoba & Jamróz, 2013; Niedoba, 2009, 2011, 2013a,b, 2014; Niedoba & Surowiak, 2012; Surowiak & Brożek, 2014a,b; Tumidajski, 1997; Tumidajski & Saramak, 2009). It is a natural thing that the analysis of multidimensional properties and statistics begin with the analysis of particle density distribution and particle size distribution of coals and then is extended on the basis of further coal properties, especially the contents of components and reactions to the processes of its processing. The analysis of coal in terms of distribution of the so-called class-fraction is the initial information on the coal capability of developing the surface area and concentration of flammable and volatile parts and ash. While multidimensional methods of visualisation allow for the combined interpretation of all measured properties in tested terms.

2. Data

In order to investigate the mineral beneficiation capability intended for the process of gasification in fluidised bed – bituminous coals from Janina Mining Plant (coal of 31.2 type) and Wieczorek Coal Mine (coal of 32 type) – each of them was subjected to the beneficiation process in the laboratory ring jig (10 rings, coal in the class of 0-18 mm). After the completion of the separation process, material was divided into 5 layers (with 2 rings) and each of them was sieved on sieves into 10 grain classes, establishing yields of layers and classes. Then, products obtained in such a way – grain classes, after the separation of analytical samples, were subjected to chemical elemental and technical analysis of coal in order to characterise features influencing gasification processes.

In total, from both coal mines, 99 samples (50 from Janina coal mine and 49 from Wieczorek coal mine – in one of the layers, 16-18 mm class was not obtained) having the following parameters were obtained: Total sulphur content, hydrogen content, nitrogen content, chlorine content, total carbon content, heat of combustion and ash content. The card of technological suitability of coal was used additionally (Sobolewski et al., 2012, 2013), in which key, relevant and additional parameters, having effect on the gasification process, were described. On the card's basis coal samples, which are subjected to the gasification process in an effective way, were identified. Conditions used are: calorific value [kJ/kg] > 18000, ash content $< 25\%$, chlorine content $< 0.1\%$, total sulphur content $< 2\%$, carbon content $> 60\%$, $3.5\% \leq$ hydrogen content $\leq 5.5\%$, nitrogen content $< 2\%$. On the basis of the presented conditions, among the analysed 99 samples only 18 samples were identified as those which can be subject to gasification in an effective way. Among those 18 samples, 17 came from Janina Mining Plant and only one sample came from Wieczorek Coal Mine. Table 1 presents an example of the obtained data.

TABLE 1

Elemental analysis of coal in layer I after beneficiation in the jig – Janina Mining Plant

Class d [mm]	Total sulphur content S_t^a [%]	Hydrogen content H^a [%]	Nitrogen content N^a [%]	Chlorine content Cl^a [%]	Total carbon content C^a [%]	Heat of combustion Q_s^a [kJ/kg]	Ash content A^a [%]
< 1.00	1.20	3.70	0.97	0.134	61.70	24 060	12.90
1.00-2.00	0.84	4.12	1.09	0.111	67.70	26 657	4.30
2.00-3.15	0.83	4.15	1.02	0.122	67.60	26 610	4.40
3.15-5.00	0.83	4.16	1.07	0.105	67.70	26 639	4.80
5.00-6.30	0.83	4.17	1.05	0.099	67.60	26 630	4.60
6.30-8.00	0.82	4.28	1.13	0.096	67.20	26 523	4.80
8.00-10.00	0.78	4.10	1.10	0.084	67.70	26 525	4.10
10.00-12.50	0.93	4.21	1.12	0.079	67.20	26 529	4.90
12.50-16.00	0.82	4.22	1.10	0.099	67.30	26 711	5.40
16.00-18.00	0.87	4.36	1.11	0.10	68.00	27 066	5.00

3. Principal component analysis

The methods of multidimensional visualisation are increasingly used instruments for statistical analysis. A number of such methods were described in many publications (Aldrich, 1998; Asimov, 1985; Assa et al., 1999; Chatterjee et al., 1993; Cleveland, 1984; Cook et al., 1995; Chou et al., 1999; Inselberg, 1985; Jain & Mao, 1992; Kim et al., 2000; Kraaijveld et al., 1995; Gennings et al., 1999; Sobol & Klein, 1989). Authors also used methods of this type for analyses and classifiications of coal type (Jamróz, 2011, 2009; Jamróz & Niedoba, 2013, 2014; Niedoba, 2013, 2014; Niedoba & Jamróz, 2013). One of the methods is PCA (*Principal Component Analysis*).

3.1. The description of the method

PCA is one of the statistical methods of factor analysis. In this method, the orthogonal projection of multidimensional data in a plane represented by specially selected vectors V_1,V_2 is performed. These are eigenvectors, corresponding to the two largest (in terms of a module) eigenvalues of the covariance matrix of the observation space. The described selection of vectors V_1,V_2 enables to obtain the image in a plane representing the most variability in the data.

3.2. The algorithm

The input data set consists of elements described by n-properties. It can be therefore treated as a set of n-dimensional vectors. Let us identify k-th input data vector as $x_k = (x_{k,1}, x_{k,2}, \dots x_{k,n})$. The algorithm serving to realise the visualisation using PCA consists of several steps:

a) Input data scaling. Individual properties, represented by individual data dimensions are scaled in such a way so as to fall into the same preset interval. It was decided to scale individual coordinates (properties) of vectors of the data set to the interval (0, 1).

b) Covariance matrix determination. We use the general formula for the covariance:

$$\mathrm{cov}(X,Y) = E(XY) - E(X) \cdot E(Y) \tag{1}$$

where E denotes the expected value. At first, we thus calculate expected values:

$$E_i = \frac{\sum_{k=1}^{m} x_{k,i}}{m} \tag{2}$$

and

$$E_{i,j} = \frac{\sum_{k=1}^{m} x_{k,i} x_{k,j}}{m} \tag{3}$$

where E_i – the expected value i-th coordinate of the input data, $E_{i,j}$ – the expected value of the quotient of i-th and j-th coordinate of the input data, m – a number of input data vectors, $x_{k,i}$ – i-th coordinate of the k-th input data vector.
If we denote the covariance matrix as A, then each element of the matrix a_{ij} is obtained by counting:

$$a_{ij} = E_{i,j} - E_i E_j \tag{4}$$

In this way, a symmetrical covariance matrix of the input data set is obtained.

c) Determination of eigenvalues and eigenvectors of the covariance matrix. For the numerical calculations, the Jacobi method was selected. In this method, we use the fact that the orthogonal transformation does not change the eigenvalues and eigenvectors of the matrix. Thus we can perform a sequence of such orthogonal transformations on matrix A in order to bring it into the diagonal form D:

$$A = W \cdot D \cdot W^T \tag{5}$$

In the diagonal matrix, there are eigenvalues on the main diagonal, while the eigenvectors corresponding to them will be noted in the matrix W columns. Matrices D and W fulfilling the equation (5) using the Jacobi method can be obtained in the following steps:

1) We assume an identity matrix of size nxn as matrix W,
2) We assume covariance matrix of size nxn calculated in point b) as matrix A,
3) We select a leading element outside the main diagonal of matrix A that is such whose value is the largest in terms of the module and does not lie on the main diagonal. We look for its position in the matrix, that is such coordinates p and q, that:

$$\forall i, j = 1,\ldots,n \text{ and } i \neq j \text{ is}: \left|a_{pq}\right| \geq \left|a_{ij}\right| \tag{6}$$

4) We calculate values c and s. At first we determine:

$$r = \frac{a_{qq} - a_{pp}}{2a_{pq}} \tag{7}$$

$$t = \frac{\operatorname{sgn}(r)}{|r| + \sqrt{r^2 + 1}} \tag{8}$$

where: a_{ij} denotes the element of the matrix from i-th row and j-th column, sgn(r) = 1 for $r >= 0$ and sgn(r) = –1 for $r < 0$. Then we determine:

$$c = \frac{1}{\sqrt{t^2 + 1}} \tag{9}$$

and

$$s = tc \tag{10}$$

5) Using calculated values c and s, we create matrix B in such a way that it is identity matrix of size nxn, in which we change four elements: $b_{pp} = c$, $b_{qq} = c$, $b_{pq} = s$, $b_{qp} = -s$,
6) We assume new value to matrix A, using the current value of matrix A, matrix B created in the previous step and transposed matrix B:

$$A = B^T \cdot A \cdot B \tag{11}$$

7) We assume new value to matrix W, using the current value of matrix W and matrix B created in step 5:

$$W = W \cdot B \tag{12}$$

8) We check if, as a result of calculations, we obtained the assumed at the beginning accuracy of calculations ε, that is:

$$\frac{\max\limits_{i,j=1,\ldots,n \wedge i \neq j} |a_{ij}|}{\max\limits_{i=1,\ldots,n} |a_{ii}|} < \varepsilon \tag{13}$$

If inequality (13) is not fulfilled, we return to step 3 and continue calculations. Otherwise, obtained matrix A is a diagonal matrix. As a result, on the main diagonal of obtained matrix A there are eigenvalues of output matrix and in columns of obtained matrix W there are eigenvectors corresponding to them.

d) Two coordinate axes determination. Among the calculated at the stage described in subpoint c vectors, we select two eigenvectors corresponding to two largest, in terms of a module, eigenvalues of the covariance matrix. We denote them as $V_1 = (v_{1,1}, v_{1,2}, \ldots v_{1,n})$, $V_2 = (v_{2,1}, v_{2,2}, \ldots v_{2,n})$. In this way we obtained two coordinate axes on which we will project all data.
e) Drawing a set of points on the screen. For each point x_k we determine its two coordinates $(\tilde{x}_{k,1}, \tilde{x}_{k,2})$ obtained after the projection onto axes V_1 and V_2, that is:

$$\tilde{x}_{k,1} = \sum_{i=1}^{n} v_{1,i} x_{k,i} \tag{14}$$

$$\tilde{x}_{k,2} = \sum_{i=1}^{n} v_{2,i} x_{k,i} \tag{15}$$

Thanks to this, we can present the image of each vector on the computer screen. This is realised through drawing a symbol, representing the class, to which the vector of data x_k corresponding to it belongs to, in the point with coordinates $(\widetilde{x}_{k,1}, \widetilde{x}_{k,2})$ on the screen. In this way, an image of multidimensional points representing different classes of coal appears on the computer screen.

4. The results of experiments

Within the study, in order to visualise seven-dimensional data describing coal samples, the computer system based on the assumptions outlined in the previous point was developed. It was written in C++ programming language with the use of Microsoft Visual Studio. The obtained results are presented in Figs. 1-3. These views show the way in which 7-dimensional data is transformed by means of PCA into two dimensions. The algorithm of visualisation by means of PCA works in this way, despite a considerable reduction in the number of dimensions, so as to obtain the view presenting the largest variability in data. In this way, we can see important properties of 7-dimensional data features on the 2-dimensional screen.

It was decided to check if the method of multidimensional data visualisation enables to divide the space of samples into areas with different suitability for the fluidised gasification process. Figures 1-3 present views of points representing seven-dimensional vectors of data describing coal samples obtained from Janina and Wieczorek coal mines. To obtain them, the developed visualisation system calculated the covariance matrix:

$$\mathrm{cov} = \begin{bmatrix} 0.0350 & -0.0202 & -0.0294 & -0.0081 & -0.0231 & -0.0224 & 0.0173 \\ -0.0202 & 0.0583 & 0.0640 & 0.0173 & 0.0621 & 0.0615 & -0.0595 \\ -0.0294 & 0.0640 & 0.0816 & 0.0254 & 0.0686 & 0.0686 & -0.0587 \\ -0.0081 & 0.0173 & 0.0254 & 0.0295 & 0.0183 & 0.0187 & -0.0130 \\ -0.0231 & 0.0621 & 0.0686 & 0.0183 & 0.0672 & 0.0664 & -0.0648 \\ -0.0224 & 0.0615 & 0.0686 & 0.0187 & 0.0664 & 0.0658 & -0.0635 \\ 0.0173 & -0.0595 & -0.0587 & -0.0130 & -0.0648 & -0.0635 & 0.0680 \end{bmatrix}$$

and eigenvectors (corresponding to the two largest in terms of module eigenvalues):

V_1 = (–0.1666 0.4115 0.4673 0.1364 0.4439 0.4395 –0.4192),

V_2 = (–0.6610 –0.1188 0.3043 0.5051 –0.1160 –0.1005 0.4215)

Figure 1 presents the illustration of the discussed data according to the division into coal samples from Janina and Wieczorek coal mines. In this figure, it is clearly visible that the images of points representing samples of coal from different coal mines occupy separate subareas and accumulate in clusters. It is clearly seen here that in the whole area of the figure, these clusters can be easily separated from each other. On the basis of this figure, it can be stated that the PCA method of multidimensional data visualisation enables to divide the space of samples into areas belonging to different coal mines. Thanks to this, by analysing the next, unknown samples we can qualify them according to their origin into a group coming from Janina or Wieczorek coal mine through their visualisation.

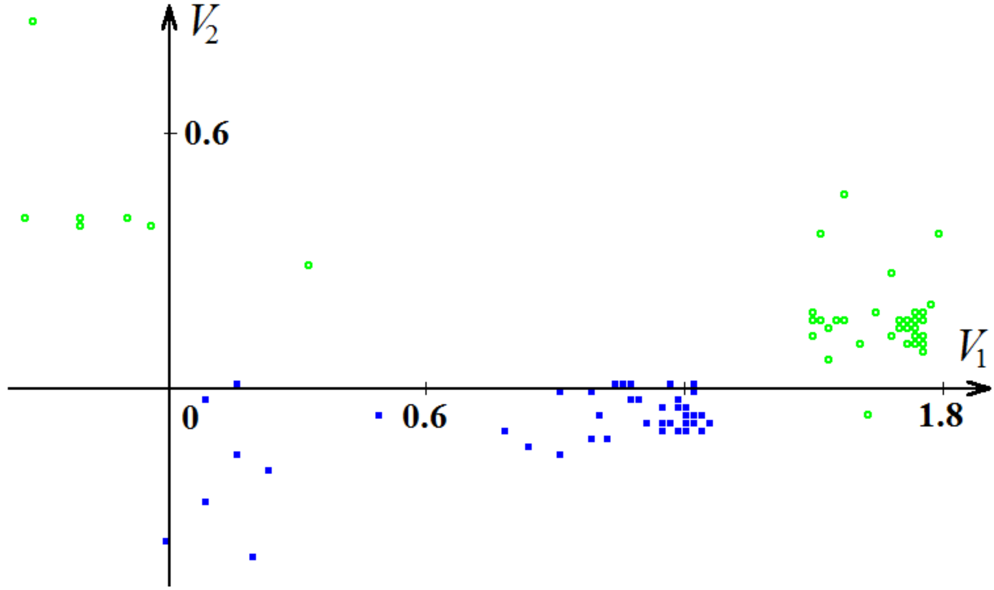

Fig. 1. The view of 7-dimensional data with the division according to the place of extraction. The images of points representing coal samples obtained in Janina coal mine are marked with a symbol of square (■), coal samples obtained in Wieczorek coal mine – marked with circle (○)

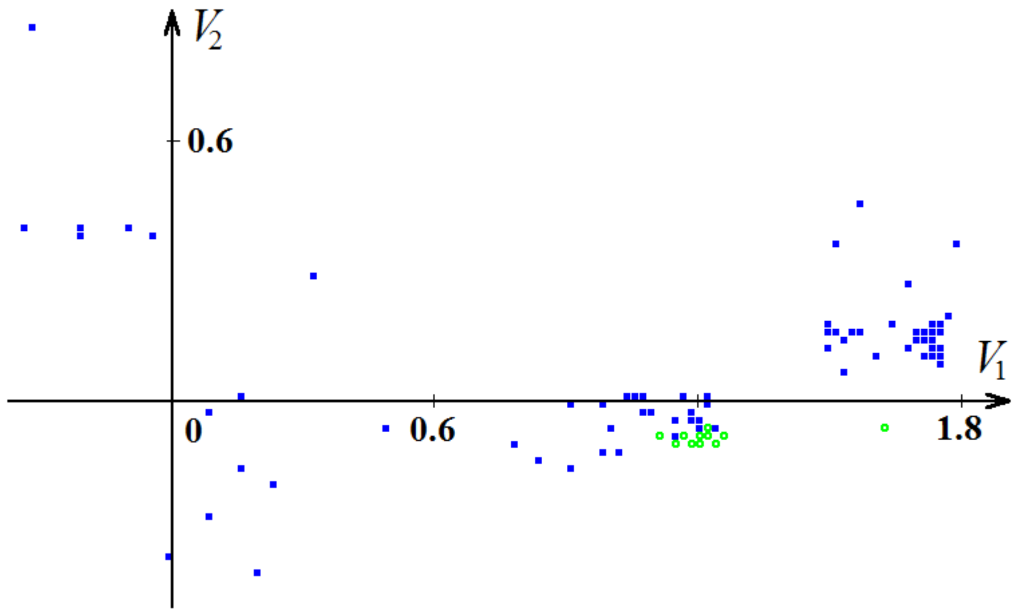

Fig. 2. The view of 7-dimensional data representing samples of coal with different suitability for the fluidised gasification process. The images of points representing coal samples less suitable for gasification are marked with a symbol of square (■), coal samples more suitable for gasification – marked with circle (○)

Figure 2 presents the discussed data according to a completely different division – the division into samples of coal more susceptible to gasification and less susceptible to gasification. In this figure, it is visible that the images of points representing samples of coal more susceptible to gasification and less susceptible to gasification occupy separate subareas and accumulate in clusters. It is seen that in the whole area of the figure, these clusters can be easily separated from each other. On the basis of this figure, it can be stated that the PCA method of multidimensional data visualisation enables to divide the space of samples into areas with different suitability for the fluidised gasification process. Thanks to this, by analysing the next, unknown samples we can qualify them into a group of more suitable samples for gasification or less suitable samples for gasification through their visualisation. This is especially important because, in the analysed situation, coal samples more suitable for gasification occupy the interior of the seven-dimensional cuboid – which is a considerable simplification. It results directly from the fact that the assumed conditions specifying belonging to this group (the card of technological suitability of coal) are simple inequalities with which you can easily check such belonging. In fact, it may, however, turn out that the area of belonging can have considerably more complicated shape. Then on the basis of larger number of samples whose belonging to the class of coal more suitable for gasification will be established empirically, it will be possible to try using PCA to obtain space division into areas representing samples of coal more and less suitable for gasification. Thanks to this, it can turn out that the obtained mapping will reflect reality more accurately. Therefore, the earlier statement that the PCA method of multidimensional data visualisation enables to divide the space of samples into areas with different suitability for the fluidised gasification process takes particular effect.

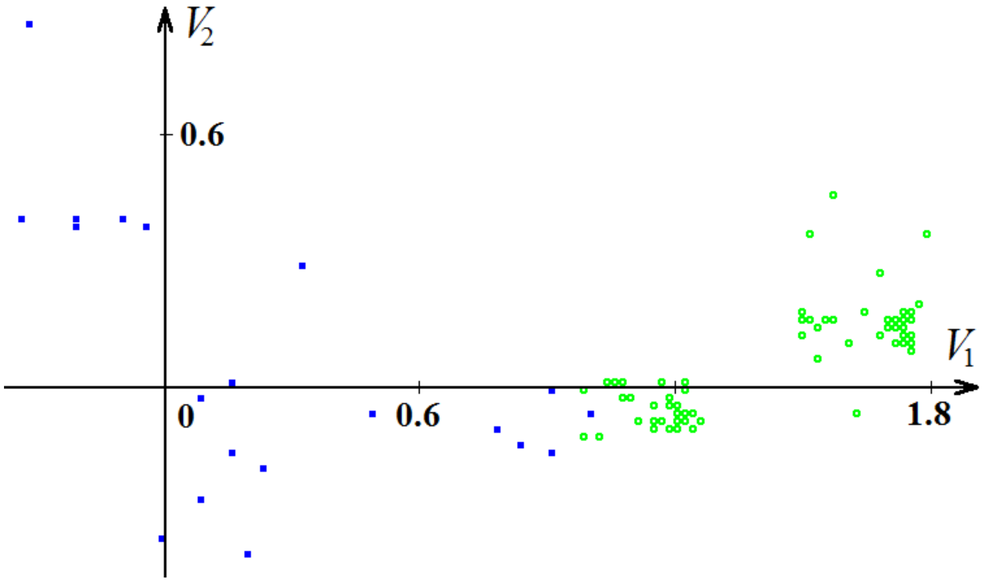

Fig. 3. The view of 7-dimensional data representing samples of coal with different suitability for the fluidised gasification process with the omission of the condition for the chlorine content. The images of points representing coal samples less suitable for gasification are marked with a symbol of square (■), coal samples more suitable for gasification – marked with circle (○)

Additionally, as it turns out, the same space division contains considerably more information. It is shown in Figure 3, in which the discussed data according to the division into samples of coal more susceptible to gasification and less susceptible to gasification with the omission of the condition for the chlorine content. Also here despite the omission of the condition for the chlorine content, the images of points representing samples of coal more and less susceptible to gasification occupy separate subareas and accumulate in clusters. It is seen that in the whole area of the figure, these clusters can be easily separated from each other. On the basis of this figure, it can be stated that the PCA method of multidimensional data visualisation enables to divide the space of samples into areas with different suitability for the fluidised gasification process even with a change in conditions determining this suitability. In this specific case, it has particular importance because the chlorine content influences only the degree of contamination resulting from gasification and not the effectiveness of this gasification. But the assignment of samples changes completely. For comparison, in Figure 2 only 18 samples were identified as those which can be subject to gasification in an effective way. Among those 18 samples, 17 came from Janina coal mine and only one sample came from Wieczorek coal mine. While with the omission of the condition for the chlorine content (Figure 3), among the analysed 99 samples of coal as much as 78 samples were identified as those which can be subject to gasification in an effective way. In this case, 78 samples, 37 came from Janina coal mine and 41 samples came from Wieczorek coal mine. It can be concluded from this that if we can omit the chlorine contamination, then the use of coal from Wieczorek coal mine for gasification will be more effective – otherwise from Janina coal mine.

It should be noted that by the visualisation algorithm through PCA does not use information on the points representing data belonging to specific classes. In this situation, the way in which the images of points representing a given class will be grouped depends only on certain, properties of this data identified by the algorithm. Therefore, Figures 1-3 differ only in the belonging of individual points to different classes. It results from the fact that all three figures were created as a result of projection of individual data vectors onto two eigenvectors corresponding to the two largest in terms of module eigenvalues of the same covariance matrix. Because for all figures the covariance matrix is calculated for exactly the same seven-dimensional data with the omission of information on the belonging of the points to individual classes. Therefore, the location of points in all three figures is identical – only their assignment to the respective classes is different.

5. Conclusions

The conducted experiments consisting of the visualisation of seven-dimensional data using PCA enabled to obtain the following conclusions:

1) the multidimensional visualisation using PCA enables to state that the images of points representing samples of coal more susceptible to gasification and less suitable for gasification occupy separate subareas of space and accumulate in clusters which can be easily separated from each other. The PCA method enables to divide the space of samples into areas with different suitability for the fluidised gasification process. Thanks to this, by analysing the next, unknown samples we can qualify them into a group of more suitable samples for gasification or less suitable samples for gasification through their visualisation.

2) as a result of the multidimensional visualisation using PCA it is possible to state that the images of points representing samples of coal from Janina and Wieczorek coal mines occupy separate subareas and accumulate in clusters which can be easily separated from each other. Thanks to this, the space of samples can be divided into areas belonging to different coal mines. Thanks to this, by analysing the next, unknown samples we can qualify them according to their origin into a group coming from Janina or Wieczorek coal mine through their visualisation.
3) the algorithm of the visualisation through PCA does not use information on the belonging of the points representing data to specific classes. In this situation, the way in which the images of points representing a given class will be grouped depends only on certain, properties of this data identified by the algorithm – irrespectively of their allocation to different classes.
4) the same division of the space of samples conducted using PCA at the same time groups the points representing the analysed data both in terms of place of their extraction (Janina and Wieczorek coal mines) and their suitability for the fluidised gasification process.
5) On the basis of the card of technological suitability of coal, among the analysed 99 samples only 18 samples were identified as those which can be subject to gasification in an effective way. Among those 18 samples, 17 came from Janina coal mine and only one sample came from Wieczorek Coal Mine.
6) The situation changes dramatically with the omission of the condition for the chlorine content. Then on the basis of the same card of technological suitability of coal, among the analysed 99 samples of coal as much as 78 samples were identified as those which can be subject to gasification in an effective way. Among those 78 samples, 37 came from Janina coal mine and 41 samples came from Wieczorek coal mine.
7) The undoubted advantage of the PCA method is the fact that during the visualisation there is no necessity to select any parameters, in contrast to many other methods of the multidimensional data visualisation.

References

Ahmed H.A.M., Drzymała J., 2005. *Two-dimensional fractal linearization of distribution curves*. Physicochemical Problems of Mineral Processing, Vol. 39, p. 129-139.

Aldrich C., 1998. *Visualization of transformed multivariate data sets with autoassociative neural networks*. Pattern Recognition Letters, Vol. 19(8), p. 749-764.

Asimov D., 1985. *The Grand Tour: A Tool for Viewing Multidimensional Data*. SIAM Journal of Scientific and Statistical Computing, Vol. 6, p. 128-143.

Assa J., Cohen-Or D., Milo T., 1999. *RMAP: a system for visualizing data in multidimensional relevance space*, Visual Computer, Vol. 15(5), p. 217-34.

Blaschke W., 2009. *Hard coal processing – gravitational beneficiation*. Wydawnictwo IGSMiE PAN, Kraków [in Polish].

Borowiecki T., Kijeński J., Mochnikowski J., Ściążko M. (ed.), 2008. *Pure energy, chemical products and coal made fuels – evaluation of development potential*. IChPW, Zabrze [in Polish].

Brożek M., Surowiak A., 2005. *The Dependence of Distribution of Settling Velocity of Spherical Particles on the Distribution of Particle Sizes and Densities*. Physicochemical Problems of Mineral Processing, Vol. 39, p. 199-210.

Brożek M., Surowiak A., 2007. *Effect of Particle Shape on Jig Separation Efficiency*. Physicochemical Problems of Mineral Processing, Vol. 41, p. 397-413.

Brożek M., Surowiak A., 2010. *Argument of Separation at Upgrading in the Jig*. Archives of Mining Sciences, Vol. 55, p. 21-40.

Chatterjee A., Das P.P., Bhattacharya S., 1993. *Visualization in linear programming using parallel coordinates*. Pattern Recognition, Vol. 26(11), p.1725-1736.

Chmielniak T., Tomaszewicz G., 2012. *Solid fuels gasification – prezent state and future development directions*. Karbo, Vol. 3, p. 191-201 [in Polish].

Chou S.-Y., Lin S.-W., Yeh C.-S., 1999. *Cluster identification with parallel coordinates*. Pattern Recognition Letters, Vol. 20, p. 565-572.

Cleveland W.S., McGill R., 1984. *The many faces of a scatterplot*. Journal of the American Statistical Association, vol.79, pp.807-822.

Cook D., Buja A., Cabrera J., Hurley C., 1995. *Grand Tour and Projection Pursuit*. Journal of Computational and Graphical Statistics, Vol. 4, no. 3, p. 155-172.

Drzymała J., 2009. *Basics of minerallurgy*. Oficyna Wydawnicza Politechniki Wrocławskiej, [in Polish].

Gennings C., Dawson K.S., Carter W.H., Jr. Myers R.H., 1990. *Interpreting plots of a multidimensional dose-response surface in a parallel coordinate system*. Biometrics, Vol. 46, p. 719-735.

Inselberg A., 2009. *Parallel Coordinates: VISUAL Multidimensional Geometry and its Applications*. Springer.

Jain A.K., Mao J., 1992. *Artificial neural network for non-linear projection of multivariate data*. In: *Proc. IEEE Internat. Joint Conf. On Neural Networks*. Baltimore, MD, Vol. 3, p. 335-340.

Jamróz D., 2001. *Visualization of objects in multidimensional spaces*. Doctoral Thesis, AGH, Kraków [in Polish].

Jamróz D., 2009. *Multidimensional labyrinth – multidimensional virtual reality*. [In:] Cyran K., Kozielski S., Peters J., Stanczyk U., Wakulicz-Deja A. (eds.), *Man-Machine, Interactions*, AISC, Heidelberg, Springer-Verlag, Vol. 59, p. 445–450.

Jamróz D., 2014a. *Application of Multidimensional Data Visualization in Creation of Pattern Recognition Systems*. [In:] Gruca A., Czachórski T., Kozielski S. (eds.), *Man-Machine, Interactions 3*, AISC, Switzerland, Springer International Publishing, Vol. 242, p. 443-450.

Jamróz D., 2014b. *Application of multidimensional scaling to classification of various types of coal*. Archives of Mining Sciences, Vol. 59(4), p. 413-425.

Jamróz D., 2014c. *Application of multi-parameter data visualization by means of autoassociative neural networks to evaluate classification possibilities of various coal types*. Physicochemical Problems of Mineral Processing, Vol. 50(2), p. 719-734.

Jamróz D., Niedoba T., 2014. *Application of Observational Tunnels Method to Select Set of Features Sufficient to Identify a Type of Coal*. Physicochemical Problems of Mineral Processing, vol 50(1), p. 185-202.

Kim S.-S., Kwon S., Cook D., 2000. *Interactive visualization of hierarchical clusters using MDS and MST*. Metrika, Springer-Verlag, Vol. 51, p. 39-51.

Kosminski A., Ross D.P., Agnew J.B., 2006. *Transformations of sodium during gasification of low-rank coal*. Fuel Processing Technology, p. 943-952.

Kraaijveld M., Mao J., Jain A.K., 1995. *A nonlinear projection method based on Kohonen's topology preserving maps*. IEEE Trans. Neural Networks, Vol. 6(3), p. 548-559.

Lee S., Speight J.G., Loyalka S.K., 2007. *Handbook of alternative fuel technologies*. CRC Press, Taylor & Francis Group.

Marciniak-Kowalska J. et al. Project report NCBiR, 2012-13: *Elaboration of technology of coal gasification for highly efficient fuel and energy production*, report from part of the project *Investigations of coal beneficiation by means of mechanical processing*, not publisher work [in Polish].

Niedoba T., 2009. *Multidimensional distributions of grained materials characteristics by means of non-parametric approximation of marginal statistical density function*, AGH Journal of Mining and Geoengineering, Vol. 4, p. 235-244 [in Polish].

Niedoba T., 2011. *Three-dimensional distribution of grained materials characteristics*. [In:] Proceedings of the XIV Balkan Mineral Processing Congress, Tuzla, Bosnia and Herzegovina, Vol. 1, p. 57-59.

Niedoba T., 2013a. *Multidimensional characteristics of random variables in description of grained materials and their separation processes.* Wydawnictwo Instytutu Gospodarki Surowcami Mineralnymi i Energią PAN, Kraków [in Polish].

Niedoba T., 2013b. *Statistical analysis of the relationship between particle size and particle density of raw coal*. Physicochemical Problems of Mineral Processing, Vol. 49(1), p. 175-188.

Niedoba T., 2014. *Multi-parameter data visualization by means of principal component analysis (PCA) in qualitative evaluation of various coal types*. Physicochemical Problems of Mineral Processing, Vol. 50(2), p. 575-589.

Niedoba T., Jamróz D., 2013. *Visualization of multidimensional data in purpose of qualitative classification of various types of coal*. Archives of Mining Sciences, Vol. 58(4), p. 1317-1333.

Niedoba T., Surowiak A., 2012. *Type of coal and multidimensional description of its composition with density and ash contents taken into consideration*. in Proceedings of the XXVI International Mineral Processing Congress, Vol. 1, p. 3844-3854.

Sobol M.G., Klein G., *New graphics as computerized displays for human information processing.* IEEE Trans. Systems Man Cybernet, Vol. 19 (4), p. 893-898, 1989.

Sobolewski A., Chmielniak T., Topolnicka T., Gieza N., 2013. *Selection of coals to gasification in pressure fluidized bed gas generator*. Karbo, Vol. 1, p. 28-38 [in Polish].

Sobolewski A., Chmielniak T., Topolnicka T., Świeca G., 2012. *Characteristics of Polish coals in aspekt of their suitability to fluidized gasification*. Polish Mining Review, Vol. 2, p. 174-183 [in Polish].

Strugała A., Czalicka-Kolarz K., Ściążko M., 2011. *Projects of new technologies of coal gasification created within Strategic Program NCBiR*. Energy Policy Journal, Vol. 14(2), p. 375-390.

Strugała A., Czerski G., 2012. *Investigations over technologies of coal gasification in Poland*, Chemical Review, Vol. 91(11), p. 2181-2185 [in Polish].

Surowiak A., 2013a. Assessment of coal mineral matter liberation efficiency index, Minerals Engineering, Vol. 14(2), p. 153-158.

Surowiak A., 2013b. *Investigations over beneficiation of hard coals destined to gasification in fluidized Bed gas generator*, Polish Mining Review, Vol. 69(2), p. 239-244 [in Polish].

Surowiak A., 2014a. *Possibilities of upgrading of hard coals destined to ground gasification process*. Polish Mining Review, Vol. 70(4), p. 59-66 [in Polish].

Surowiak A., 2014b. *Influence of particle density distributions of their settling velocity for narrow size fractions*. Mineral Resources Management, Vol. 30(1), p. 105-122.

Surowiak A., Brożek M., 2014a. *Methodology of calculation the terminal settling velocity distribution of spherical particles for high values of the Reynold's number*. Archives of Mining Sciences, Vol. 59(1), p. 269-282.

Surowiak A., Brożek M., 2014b. *Methodology of calculation the terminal settling velocity distribution of irregular particles for high values of the Reynold's number*. Archives of Mining Sciences, Vol. 59(2), p. 553-562.

Tumidajski T., 1997. *Stochastic analysis of grained materials properties and their separation processes*. Wydawnictwo AGH, Kraków [in Polish].

Tumidajski T., Saramak D., 2009. *Methods and models of mathematical statistics in mineral processing*. Wydawnictwo AGH, Kraków [in Polish].