

Multidimensional statistical and visualization methods in description of grained materials

Tomasz Niedoba^{1,a}, Paulina Pieta¹, Agnieszka Surowiak¹, and Dariusz Jamroz²

¹AGH - University of Science and Technology, Faculty of Mining and Geoengineering, Department of Environmental Engineering and Mineral Processing Cracow, Poland

²AGH – University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, Department of Computer Science

Abstract. As far as coal is concerned, the data are usually considered to be independent variables, and such an approach is not always appropriate. Therefore, the paper focuses on the multidimensional analysis, which allows to conduct the comparisons of the coal types and to determine the relationship between their specific characteristics. The paper presents an analysis of variance and an observational tunnels method, which enabled to examine the differences between three types of coals: 31, 34.2 and 35. In order to achieve the experimental aim, a number of laboratory analyses of coal were carried out. Such parameters as combustion heat, ash contents, sulfur contents, volatile parts contents and analytical moisture were determined for samples, including the mass and density of these fractions, so seven various features were specified for each coal. The results of the observational tunnels method and the ANOVA application confirmed that the accepted parameters were sufficient for the proper identification of the coal sample origin and their size fractions. As a result of ANOVA, the volatile matter content was determined as the feature on which identification of the coal types can be based, regardless of size fraction. However, the remaining parameters affect the study groups randomly. While applying the methods of multidimensional analysis, some limitations were encountered that indicated the complex structure of the grained material.

1. Introduction

Mineral resources are characterized by different physical, chemical and mechanical properties. This fact determines their complex nature which can be described by a multidimensional vector of grain property set. The grained material is a set of random variables, and this property makes it possible to carry out a multidimensional analysis, which allows a more complete examination of objects and material characteristics. There is a tendency for a comprehensive study of various phenomena and materials to be carried out in order to explore thoroughly a number of factors affecting them. As it is shown in the paper, the multidimensional statistical and visualization methods are appropriate to consider the problems associated with minerals engineering. The examples described in this paper confirmed the high potential that they carry. Consideration of grained materials classification and

^aCorresponding author: tniedoba@agh.edu.pl

identification constitutes only one of many issues which can be faced in the field of mineral processing. A number of works serve as an evidence of the widespread use of the analysis of variance and the observational tunnels method in different areas of science [4, 10, 11, 14, 16-18]. However, the analysis carried out for different types of coal confirms that these methods can be successfully connected to the mineral processing. The methods of multidimensional data visualization have their origins in statistics and computer science as well as they allow for graphical interpretation of large amounts of data. The multidimensional data analyses are complicated, but they are conducted more frequently in the era of computerization [15, 21].

2. Experiment

In order to introduce the topic of the paper, a number of laboratory analyses of coal were carried out. Coals classified, according to the Polish classification, as types 31, 34.2 and 35 came from some mines located within the Upper Silesian Industrial Region. At the beginning, an analysis of particle size distribution of each coal type was screened on a set of sieves of following sizes: -1.00, -3.15, -6.30, -8.00, -10.00, -12.50, -14.00, -16.00 and -20.00 mm. After that the density analysis was performed in zinc chloride suspensions of various densities (1.3, 1.4, 1.5, 1.6, 1.7, 1.8 and 1.9 g/cm³). The density fractions were determined by the following characteristics: particle size, density, ash content, volatile matter content, sulfur content, heat of combustion and analytical moisture. It was assumed that the resulting measured values were continuous random variables [17]. The results of laboratory analyses allowed to carry out the one-dimensional analysis of variance and the visualization of multidimensional data by the observational tunnels method. Energy and coking coals can be distinguished in classification of coals types. The coal properties are usually considered as independent variables, and such an approach is not always appropriate. Therefore, the paper focuses on the multidimensional analysis, which allows to conduct the comparisons of the coal types and to determine the relationship between specific characteristics of this energy source.

3. Results and discussion

3.1. Visualization of 7-dimensional data by observational tunnels method

A large amount of experimental data concerning an object can bring particular difficulties in their interpretation and reasoning. On the other hand, separate examination of variables affecting the phenomenon will limit a spectrum of relationship occurring among them in reality. A qualitative multidimensional analysis can therefore be applied using methods of data visualization, which change multidimensional features of either phenomena or material vectors into a form in which it is possible to transfer data onto either a two- or three-dimensional plane to examine certain areas of homogeneity. The results of these analyses can be helpful in characterizing grained materials, as well as in constructing mineral processing models [18]. Up until now, the methods applied to provide a visual image of multidimensional data were used to examine different problems. Those methods are: grand-tour [6], principal component analysis [12], neural networks [7, 13], parallel coordinates [3], scatter-plot matrices method [2] and mosaic plots [5]. In the paper, the results of another multidimensional visualization of data are presented [8, 9]. The observational tunnels method makes it possible to achieve an external view of the observed multidimensional sets of points using the tunnel radius. [18].

Methodology of the observational tunnels method was described elsewhere [8]. It can be stated that the method used the parallel and local orthogonal projections of an extent limited by the maximal radius of the tunnel. This connection made it possible to observe selected parts of a space in order to obtain important information on an object. Applying only one of these methods does not provide such opportunities. The method of projection used in this paper is presented in a demonstrative manner in Fig. 1 [8, 10, 11, 16, 17].

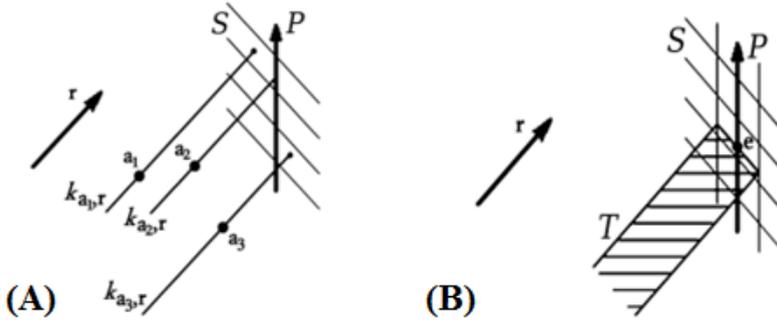


Figure 1. Presentation of projection rules on plane P in observational tunnels method (A) and the way of choosing observational tunnel T (B). Source: [18].

Mathematical tools used to describe the space, in which the objects will be placed, are: observation point a , observed space X , observational plane $P = \delta(w, \{p_1, p_2\})$ (1), hypersurface $S_{(s,d)}$ (2), line parallel to the r and passing through a called the set $k_{a,r}$ (3), tunnel radius b_a (4), projection of an observation point $e \in P \cap k_{a,r}$. The tools are defined as follows:

$$\delta(w, \{p_1, p_2\}) = \{x \in X: \exists \beta_1, \beta_2 \in F, \text{ such that } x = w + \beta_1 \cdot p_1 + \beta_2 \cdot p_2\} \quad (1)$$

$$S_{(s,d)} = \{x \in X: (x - s, d) = 0\} \quad (2)$$

$$k_{a,r} = \{x \in X: \exists \psi \in F, \text{ such that } x = \psi \cdot r + a\} \quad (3)$$

$$b_a = \psi \cdot r + a - w - \beta_1 \cdot p_1 - \beta_2 \cdot p_2 \quad (4)$$

where:

$$\psi = \frac{(w-a,r)}{(r,r)} \quad (5)$$

$$\beta_1 = \frac{(\psi r + a - w, p_1)}{(p_1, p_1)} \quad (6)$$

$$\beta_2 = \frac{(\psi r + a - w, p_2)}{(p_2, p_2)} \quad (7)$$

$r \in X$ - a proper direction of protection onto observational plane P ,

$\{\beta_1, \beta_2\}$ - a position of the projection of observed point a ,

$\beta_1, \beta_2 \in F$, such that $e = w + \beta_1 p_1 + \beta_2 p_2$,

$\psi \in F$, such that $e = \psi r + a$ - the distance projection of the observed point a .

The methodology of the observational tunnels method and the drawing procedure are accurately described elsewhere [8-11, 18]. The observational tunnels method was used to analyze the data describing three types of hard coal. This method served to visualize the multidimensional data, in which each of seven parameters was treated as a separate dimension, where the 7-dimensional space was created. While analyzing simultaneously all the features of coals, such views were received which did not permit the indication of a clear differentiation between different types (Fig. 2). It means that the analyzed data are complicated. Moreover, the observation tunnels method does not allow to provide constructive conclusions on the identification of size and density fractions regarding the data set [9-11]. Only individual analysis of the size fractions in pairs allowed their identification. The views obtained through the analysis confirmed that the assumed characteristics are sufficient to

identify the correct size fractions, for which the coal type originates. It was concluded on the basis of separate areas in the space occupied by the considered size fractions, which was not obtained in examination of all 9 size fractions at the same time. A similar analysis was performed for all possible pairs, only confirming the aforementioned conclusions [9-11, 18]. Therefore, the following charts are presented for the separated size fraction (1-3,15 mm).

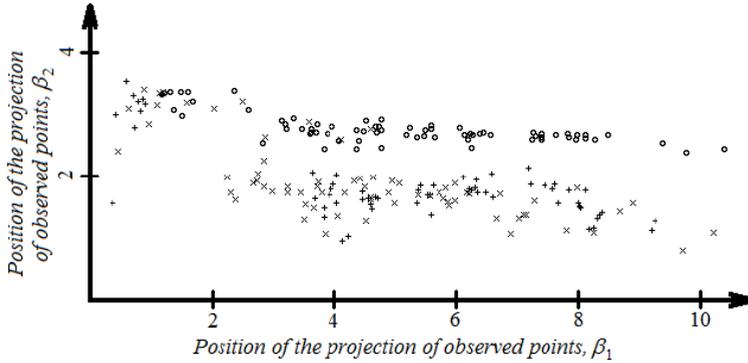


Figure 2. View of 7-dimensional data for coal 35 ("o") gather in other part of the space than coal 34.2 ("+"). From this view it is not possible to conclude about possible separation of coal of type 31 ("x"). Source: [10].

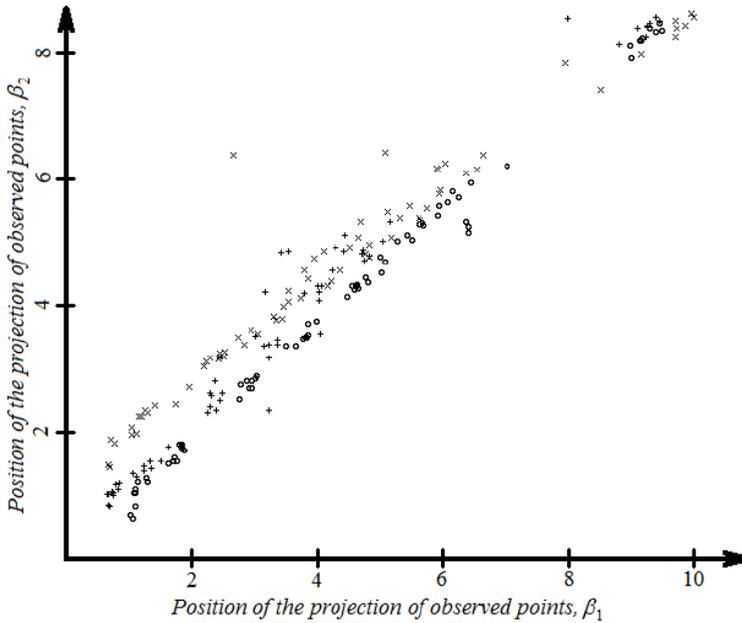


Figure 3. View of 6-dimensional set of data created after removal of parameter analytical moisture. It is visible that the data representing coal of type 35 ("o") is located in other part of the space than data representing coal type 31 ("x"). Source: [10].

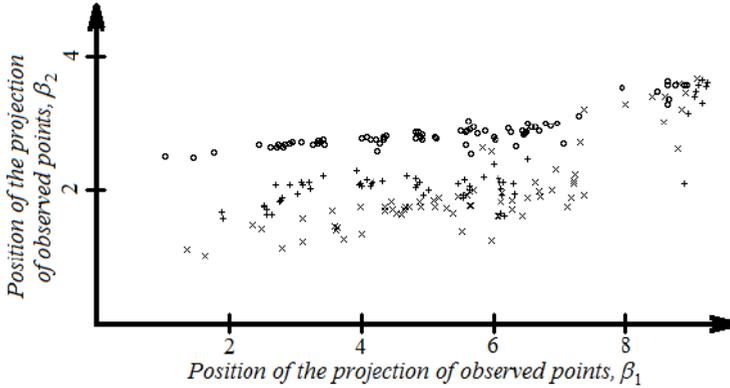


Figure 4. View of 6-dimensional data created after the removal of parameter analytical moisture. It is visible that the data representing coal type 35 ("o") is located in other part of the space than data representing coal type 34.2 ("+"). Source: [10].

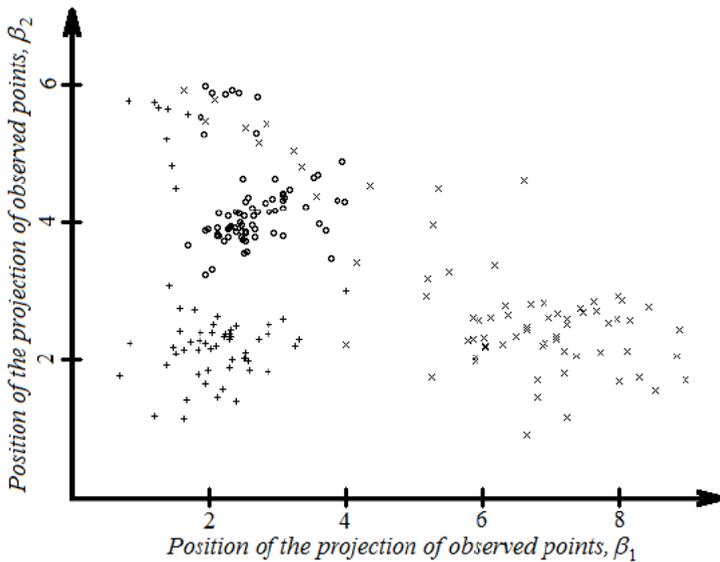


Figure 5. View of 3-dimensional data created from features: analytical moisture, sulfur contents and volatile parts contents. It is visible that the data representing coal type 34.2 ("+") is located in other part of the space than data representing coal type 31 ("x") as well data representing coal type 34.2 ("+") is located in other part of the space than data representing coal type 35 ("o"). Source: [10].

The accepted properties of coal were considered by the correlation analysis. Based on this analysis, it was stated that the influence of analytical moisture on the differentiation of coal types was small. Therefore, it was examined how its omission affected the results of the observational tunnels method (Figs. 2-4). As it was noticed before, the seven-dimensional data created from the seven coal features described above were sufficient to the proper identification of coal type, but the same data were not sufficient for this purpose after the removal of the parameter analytical moisture [10-11, 18].

The results of the correlation analysis showed that the most correlated coal features were density, combustion heat, ash contents and volatile parts contents. It can be presumed that these parameters were the most significant for coal type identification [10]. Based on correlation matrix results, the three-dimensional space was created for the mentioned features. On the basis of Figs. 6-7 it can be

concluded that the accepted characteristics were sufficient for the proper identification if the certain sample originated from coal type 31 and 35, 31 and 34.5 or 34.2 and 35.

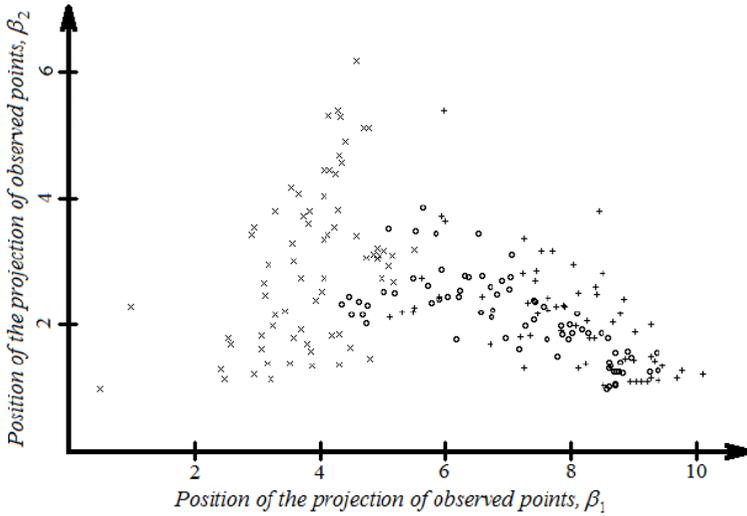


Figure 6. View of 3-dimensional data created from features: analytical moisture, sulfur contents and combustion heat. It is visible that data representing type coal 31 ("x") is located in other part of the space than data representing coal type 34.2 ("+") and other part of the space than data representing coal type 35 ("o").

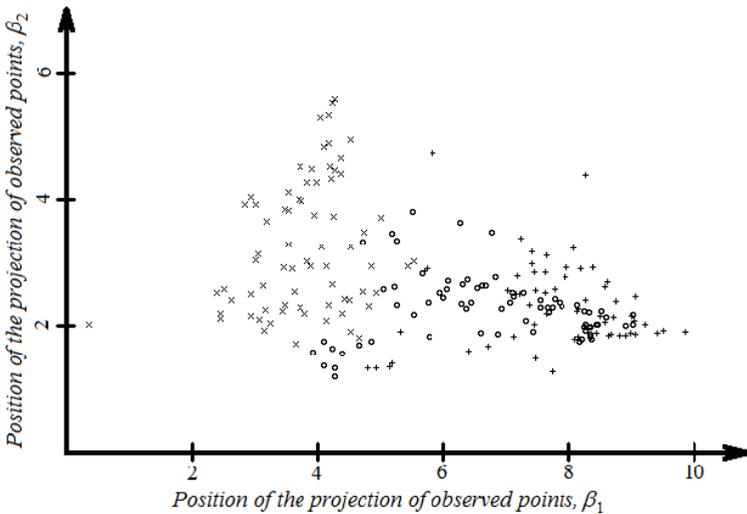


Figure 7. View of 3-dimensional data created from features: analytical moisture, sulfur contents and ash contents. It is visible that data representing coal type 31 ("x") is located in other part of the space than data concerning coal type 34.2 ("+") and other part of the space than data representing coal type 35 ("o").

At the same time it was impossible to get one view from which such conclusion could be drawn. It can be the proof that the structure of analyzed data is complicated [10-11, 18]. The next attempt was aimed to verify the influence of potential replacement of one of the coal features with another one highly correlating with it. They were characteristics highly related to each other. That is why in a set of three coal properties (analytical moisture, sulfur contents and volatile parts contents), which were

sufficient for the correct identification of coal type as shown above, the parameter volatile parts contents was replaced with the combustion heat and ash contents (Figs. 5-6). This proved that the replacement of the volatile parts contents with a mentioned features also resulted in a loss of information allowing the proper identification of all three coal types despite the high value of correlation index between the replaced coal properties [10]. It can be deduced that the volatile matter content is an essential characteristic of identifying coal types.

3.2. Examination of hard coals diversity by applying the analysis of variance

Experimental studies of certain phenomena, which are influenced by many factors, can successfully be subjected to a multidimensional parametric analysis. Variability of measurements and observations of the same objects is a widespread phenomenon which results from a number of random factors occurring during an experiment, which are simultaneously able to disturb the value of parameters. The problem of diversity was elaborated by Fisher, who developed the analysis of variance. The analysis of variance is a parametric statistical analysis, so the basis to carry out the ANOVA is to verify the assumptions concerning the homogeneity of variance and the normal distribution of these variables. For the one-dimensional analysis, the null-hypothesis assumes that the mean values μ_i of the observed characteristics in each group are equal ($H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$). The alternative hypothesis H_1 ($H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_n$) is opposite to the assumption of H_0 . In order to check the impact of parameters on a phenomenon it is necessary to calculate a value of F according to formula (8) and to determine the critical value F_α read from the F -distribution tables [19]:

$$F = \frac{\frac{\sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2}{a-1}}{\frac{\sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n-a}} = \frac{n-a}{a-1} \cdot \frac{\sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \quad (8)$$

where: n_i - the number of random variables of the studied properties,
 \bar{x} - the general average for all observations,
 \bar{x}_i - the arithmetic average of a characteristic in the studied group,
 x_{ij} - the value of a single property measurement.
 n - the number of independent observations x_{ij} equal $n = \sum_{i=1}^a n_i$
 a - the number of independent variables.

The methodology of verifying the assumption and description of different ANOVA schemes is accurately described elsewhere [20]. In the next part of this article the experiment is presented, in which ANOVA is successfully applied to study the characteristics of grained material.

A grained material has many properties that determine its classification. Such features are: content and a liberation efficiency index of valuable minerals, particle size, particle density, physicochemical properties and other physical and geometrical characteristics [1, 17, 22, 23]. In this case the ANOVA made it possible to determine main properties of coal, which differentiated the aforementioned coal types, and a comprehensive study enabled a thorough characterization of variability properties of the tested grained material. According to the scheme of the ANOVA application the fundamental assumptions were verified by Shapiro-Wilk and Brown-Forsythe tests. It allowed to exclude two properties (sulfur content, moisture analysis) from further analysis, which breached the assumption of homogeneity of variance. The F test was conducted for four features that is density, heat of combustion, ash content and volatile matter content. The results are presented in Table 1.

The F test allowed to classify the characteristics influencing the differentiation of coal types taking into account size fractions. Looking at the overall results it turned out that only the volatile matter content was a feature differentiating all tested coals in every size fraction (Fig. 8D). It can be therefore concluded that the value of this parameter was the major factor by which coal was classified according to its properties and use. This assumption indicated that a completed identification of coal types, taking into account size fractions date, was complicated and possible only for the volatile matter

content. A more detailed analysis was possible thanks to the drawn graphs showing the distribution of features values in the size fractions. The 31, 34.2 and 35 coal types were significantly different from each other in size fractions: 3.15-6.3 8-10, 10-12.5, 16-20 mm regarding four parameters. Identification of the coal types in the aforementioned size fraction taken separately could theoretically be based on four characteristics (density, heat of combustion, ash content and volatiles matter content), as indicated by *F* test results and its graphic interpretation. However, the complex structure of the analyzed data did not allow identification of hard coal types with an overall analysis of the observed characteristics in all size fraction. This was caused by the occurrence of some interactions between the values of features, which were depicted in the graphs (Fig. 8).

Table 1. Results of *F* test (significant results have been highlighted).

Size fraction [mm]	The values of empirical <i>F</i> statistics for the studied characteristics			
	Density	Heat of combustion	Ash content	Volatile matter content
0.5-1	1.466594	5.487025	2.467011	44.13694
1-3.15	1.134825	0.073532	1.815244	3.441177
3.15-6.3	4.583223	6.009332	6.946015	73.70378
6.3-8	2.399904	13.15684	4.973166	20.0778
8-10	6.414866	10.15947	7.465963	15.53309
10-12.5	6.549542	8.846621	7.088464	14.30923
12.5-14	0.335797	2.175174	2.395407	6.058163
14-16	7.623756	24.66782	0.098286	14.31082
16-20	12.08655	12.44888	10.54814	15.00584
The verification of assumptions	$F_{\alpha}(2,97)=3,0903$			

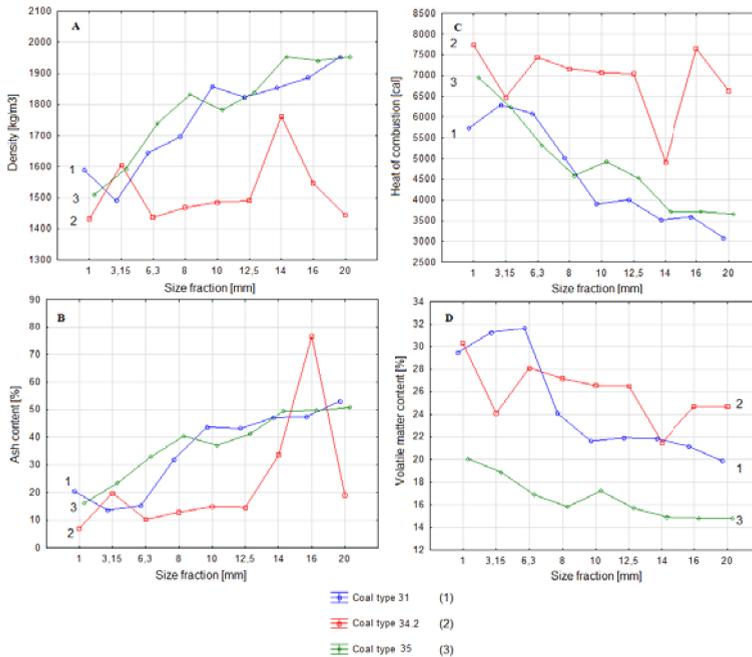


Figure 8. Graphs of weighted mean parameters (A - density, B - ash content, C - heat of combustion, D - volatile matter content) of 31, 34.2, 35 coal types.

4. Conclusions

A multidimensional description of grained material characteristics was used for evaluation of the differentiation and classification of three types of coal. This choice was dictated by the adjustment of methodology to the nature of the analyzed raw materials, which had different characteristics.

The results of the observational tunnels method application confirmed that the accepted parameters were sufficient for the proper identification of the coal sample origin and size fractions. A complicated structure of the analyzed data determined the fact that the recognition of coals or size fraction was only possible while comparing them in pairs. The visualizations of six-dimensional data made it impossible to identify the coal type. They only allowed to recognize coal types 35 and 31 as well 35 and 34.2 in pairs. Despite of the low value of the correlation coefficient for the analytical moisture, this property turned out the crucial significance for the classification of coal types. However, three-dimensional data, taking into account only the most significant characteristics (according to the correlation analysis) were enough for the proper identification of coals and the volatile matter content was necessary for appropriate recognition of coal origin.

As a result of one-dimensional analysis of variance, the volatile matter content was determined as the feature on which the identification of the coal types can be based, regardless of the size fraction. However, the remaining parameters (density, heat of combustion, ash content) affected the study groups randomly. Nevertheless, analyzing the types of coal separately in size fraction: 3.15-6.3, 8-10, 10-12.5, 16-20 mm, it can be stated that their identification was possible on the basis of all the analyzed features. In case of applying the ANOVA and the multidimensional data visualization, some limitations were encountered that indicated the complex structure of the grained material.

The results of the ANOVA and the observational tunnels method allow to conclude that the data structure is complicated. In both cases it is possible to recognize the coals in the selected size fraction on the basis of the four most important parameters: density, heat of combustion, ash content and volatiles matter content. The volatile matter content is a crucial feature that allows to classify coals according to their origin.

Acknowledgement

The paper is the effect of the scientific project no. 15.11.100.021.

References

1. Brozek M., Surowiak A., *Arch. Min. Sci.*, **55** (1), 39 (2010)
2. Eick S.G., Wills G.J., *EJOR*, **81** (3), 445 (1995)
3. Gennings C., Dawson K.S., Carter W.H., Jr. Myers R.H. *Biom.*, **46**, 719 (1990)
4. Güler T. Akdemir U., *Trans. Nonferrous Met. Soc., China*, **22**, 199 (2012)
5. Heike H., *Metrika* **51**, 11 (2000)
6. Hurley C., Buja A., *SIAM*, **11** (6), 1193 (1990)
7. Jain A.K., Mao J., in: *Proc. IEEE Internat. Joint Conf. On Neural Networks*, Baltimore, MD, **3**, 35 (1992)
8. Jamroz D., Ph.D. Thesis, AGH, University of Science and Technology, Cracow, Poland (2001)
9. Jamroz D., *Arch. Min. Sci.*, **59** (2), 413 (2014)
10. Jamroz D., Niedoba T., *Physicochem. Probl. Miner. Process.*, **50** (1), 185 (2014)
11. Jamroz D., Niedoba T., *Physicochem. Probl. Miner. Process.*, **51** (2), 769 (2015)
12. Li W., Yue H.H., Valle-Cervantes S., Qin S.J., *J Process Contr.*, **10** (5), 471 (2000)
13. Mao J., Jain A.K., *IEEE Trans. Neural Networks*, **6** (2), 296 (1995)
14. Nakhaei F. Irannajad M., Sam A., Jamalzadeh A., *Physicochem. Probl. Miner. Process.*, **52** (1), 252 (2015)

15. Niedoba T., in *proceedings of the XIV Balkan Mineral Processing Congress*, Tuzla, Bosnia and Herzegovina, **1**, 57 (2011)
16. Niedoba T., *Physicochem. Probl. Miner. Process.*, **49** (1), 175 (2013)
17. Niedoba T., *Mineral Resources Management*, Cracow, (2013)
18. Niedoba T., Jamroz D., *Arch. Min. Sci.*, **58** (4), 1317 (2013)
19. Pieta P., *5th EYEC*, Warsaw, Poland, 5, 198 (2015)
20. Stanisiz A., *StatSoft*, Cracow (2006)
21. Tumidajski T., Saramak D., AGH, Cracow (2009)
22. Tumidajski T., *MRM*, **23** (6), 111 (2010)
23. Tumidajski T. Niedoba T., Saramak D., *AGH J. of Mining and Geoenigi.*, **36** (4), 167 (2012)