



Analysis of greenhouse gas emissions in the European Union member states with the use of an agglomeration algorithm

Anna Kijewska, Anna Bluszcz*

The Silesian University of Technology, Faculty of Mining and Geology, 44-100 Gliwice, ul. Akademicka 2, Poland



ARTICLE INFO

Article history:

Received 4 November 2016

Accepted 27 February 2017

Available online 6 March 2017

Keywords:

Sustainable development

Mineral resources

Energy

Greenhouse gas emissions

Taxonomy

Cluster analysis

Agglomerative method

ABSTRACT

The use of fossil fuels as sources of energy is related to the emission of pollutants into the atmosphere. The implementation of international commitments on reducing emissions requires their continuous monitoring. The main energy resources for electricity production in the world include fossil fuels, i.e. oil, coal and natural gas, and according to projections their dominant role in the market of energy resources will persist for at least the next two decades. The aim of this article is to analyse the level of differentiation of European Union member states in terms of emissions of four greenhouse gases and to identify groups of similar countries based on these criteria. Such studies will provide information that will enrich our knowledge about the contribution of each European Union country to the emissions of greenhouse gases. This article uses a taxonomic method – cluster analysis, namely the agglomerative algorithm, which enables the extraction of objects that are similar to each other from the data and then to merge them into groups. In this way, a number of homogeneous subsets can be obtained from one heterogeneous set of objects. European Union countries make up the objects of segmentation. Each of them are described by their level of greenhouse gas emissions, such as carbon dioxide, methane, nitrogen oxides and nitrous oxides. Groups of homogeneous countries are distinguished due to total emissions and due to the level of their emissions per capita. Analysis is based on annual Eurostat reports concerning greenhouse gas emissions.

© 2017 Central Mining Institute in Katowice. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The concept of sustainable development is now having increasing impact on the economic situation of European Union countries. The assumptions of sustainable development relate to the combined economic, social and environmental spheres, and are directed at the maintenance and stimulation of economic growth whilst taking into consideration social welfare, as well as the quality of the environment as being of utmost importance. Unfortunately, for several decades, a sharp increase in greenhouse gas (GHG) emissions throughout the world has been observed. According to the Kyoto Protocol, greenhouse gases include seven gases: carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), and four fluorinated gases (F-gases). Different GHGs stay in the atmosphere for different lengths of time. Carbon dioxide is the most common GHG emitted by human activities. In 2014, in EU

countries CO₂ emission accounted for 81% of total emissions, methane for 10.6%, N₂O for 5.6%, and F-gases for 2.9% (EEA, 2016). The main emission sources in EU countries are: fuel combustion 55.1%, transport 23.2%, industrial processes and product use 8.5%, agriculture 9.9%, and waste management 3.3% (Eurostat, 2016).

A comparison of emission sources in 1990 and 2014 is shown in Fig. 1.

The use of fossil fuels as sources of energy is related to the emission of pollutants into the atmosphere. The implementation of international commitments on reducing emissions requires their continuous monitoring.

To meet its commitments to the 2005 Kyoto Protocol, the European Union created a system of measurement and limits for emissions of GHGs (UNFCCC, 2008). With the objective of reducing emissions of GHGs, the EU introduced three flexible systems: emission trading system (ETS), joint implementations (JI) and clean development mechanism (CDM) (Ranisz, 2008). The EU Emissions Trading System (EU ETS) is Europe's flagship tool to meet its carbon mitigation objectives. It remains the largest example of emissions trading in operation today, encompassing over 11 500 installations

* Corresponding author.

E-mail addresses: anna.kijewska@polsl.pl (A. Kijewska), anna.bluszcz@polsl.pl (A. Bluszcz).

across 30 countries and covering approximately 40% of total EU emissions. The three most challenging key areas of evaluation are emissions abatement in relation to the balance with economic objectives, investment and innovation impacts, and profits and price impacts. These are presented in (Laing, Sato, Grubb & Comberti, 2013).

The dynamic development of the economy and the growth of the human population is closely related to the continuing growth in demand for electricity. Oil and coal represent a significant share in the global supply of primary energy sources, as in 2014 they contributed 31% and 29% respectively. The emission of CO₂ from coal is greater than the emission from other fuels such as oil or natural gas, and accounted for 46% of the global CO₂ emissions, although coal only represented 29% of the world TPES (Total primary energy supply) in 2014 (Fig. 2). For Poland, the mining industry is a strategic sector of the economy and coal accounts for nearly 90% of electricity generation and for many years it has remained the primary source of energy in Poland (Dubiński & Turek, 2014).

Discussions are continuously ongoing regarding the impact of GHG emissions from human activities, in particular CO₂ emissions, that are dangerous for climate change. Since the Industrial Revolution, the annual CO₂ emissions from fuel combustion increased from approximately 40 Mt of CO₂ in 1813 to more than 35.85 Gt of CO₂ in 2013 (Fig. 3). The Intergovernmental Panel on Climate Change (IPCC) concluded that most of the observed increase in global average temperatures since the mid-twentieth century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations (Pachauri & Meyer, 2014). Different opinions are presented by the Nongovernmental International Panel on Climate Change (NIPCC), which represent arguments that it is nature, not human activity that governs climate (Idso, Carter, Singer, & Soon, 2013).

Currently, the subject of GHG emissions is very important because the policy of the European Union and electricity sector laws, including those relevant to the mining industry, are focused on implementing the strategy of sustainable development, mainly through the development of technologies which use renewable energy resources and the development of associated heat and

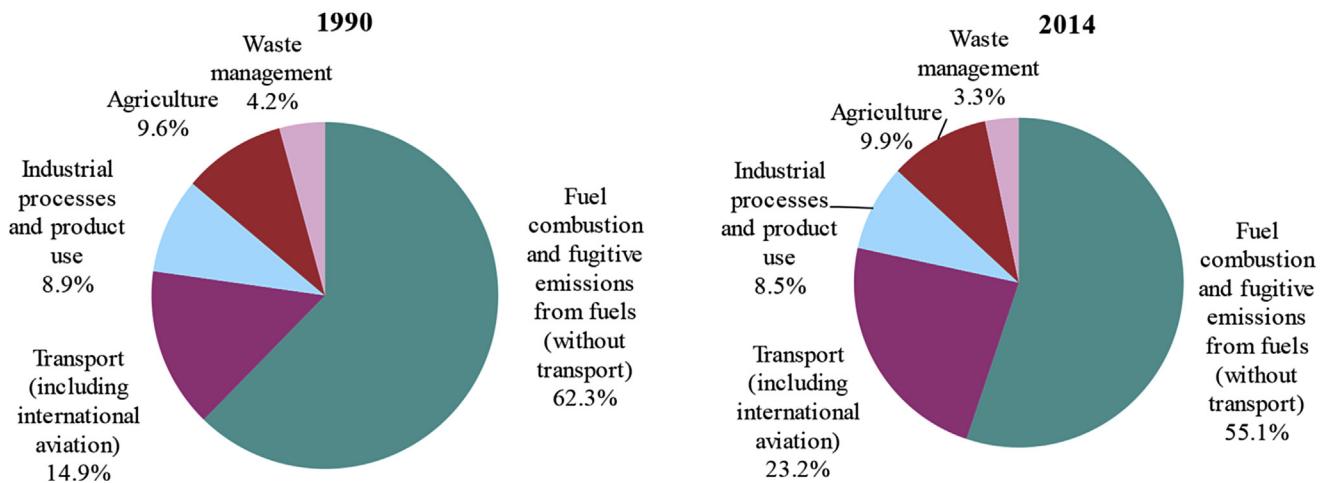


Fig. 1. The share of emissions by type of activity in 1990 and 2014 (EEA, 2016).

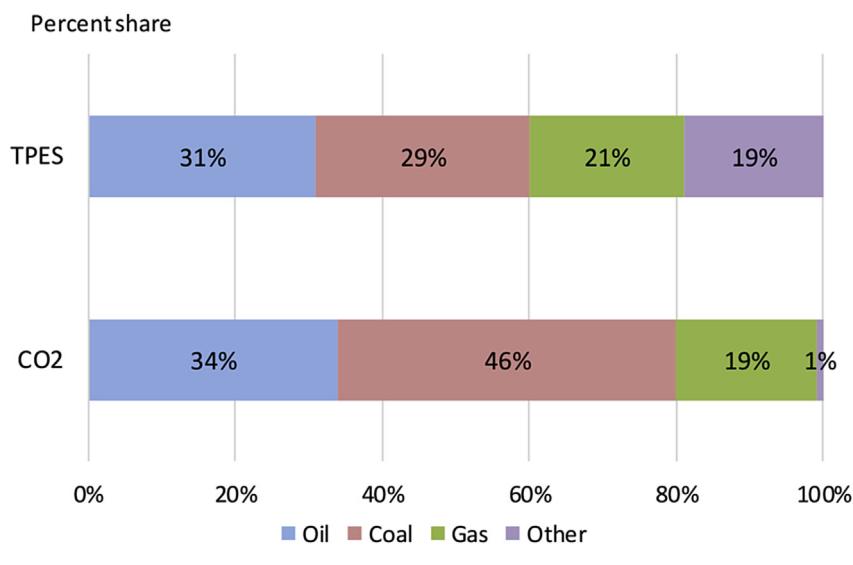


Fig. 2. World primary energy supply and CO₂ emissions: share of fuel in 2014 (IEA, 2016).



Fig. 3. Trend in CO₂ emissions from fossil fuel combustion (CDIAC, 2016).

electricity production. Sustainable energy policy is aimed at improving the long term welfare of society by striving to maintain a balance between energy security, satisfying social needs, economic competitiveness, and environmental protection (Lorek, 2011).

In 1994 the United Nations Convention made a commitment to Poland to develop and implement a national strategy for reducing greenhouse gas emissions, including economic and administrative mechanisms and the periodical monitoring of its implementation, as indicated in the "Climate Policy of Poland" (2003). Coal mining is a strategic sector of the Polish economy and plays a key role in ensuring the energy security of the country, however, it requires continuous restructuring, in order to adjust itself to market conditions (Karbownik & Stachowicz, 1994; Karbownik, Turek, & Pawełczyk, 2001; Bluszcz & Kijewska, 2015; Jonek-Kowalska, 2014; 2015; Korski, Tobór-Osadnik, & Wyganowska, 2015; 2016) and also requires the development of innovative technologies of coal combustion, which may have an impact on the reduction of the level of GHG emissions (Krzemień, Więckol-Ryk, Duda & Koteras, 2013; Krause, Krzemień & Smoliński, 2015; Lutynski, 2014; Lutynski, Sakiewicz, & Gonzalez, 2014; Stańczyk et al., 2010a; 2010b; 2012; Wiatowski et al., 2012).

At the beginning of March 2011, the European Commission adopted an action plan, which aims to reduce GHG emissions by 85–95% by 2050 in comparison to 1990. The energy policy model proposed by the European Commission aims to reduce CO₂ emissions by 40% and 60% respectively by 2030 and 2040. The ongoing policy corresponds with the internal emission reduction of 30% in 2030 and 40% in 2050 (Ciepiela, 2011). However, the predictions made by Korban and Manowska (2011) indicate that the directives of the EU to reduce CO₂ emissions by 20% compared to 1990 emissions will not be met. Real and projected emission reductions in EU countries are shown in Fig. 4.

Greenhouse gas emissions in the European Union since 1990 (taken as the base year) was reduced by 24.4% in 2014, i.e. about 1383.4 million tonnes of CO₂ eqv. The largest emitters of greenhouse gases in the European Union are: Germany, Great Britain, France, Italy, Poland and Spain. The combined share of these six countries of EU gas emissions in 2014 was over 70%. On nearly 95% of the EU's emissions 16 of the 28 member states has an impact (Fig. 5). Among these countries, the most significant decrease in emissions from 1990 to 2014 was achieved by Romania (more than 56%) and the Czech Republic (nearly 37%), while there has been an increase of emissions in three countries: Spain (15%), Portugal (6.4%) and Ireland (3.7%).

It should be noted, however, that the level of global CO₂ emissions is currently to the greatest extent affected by countries such as China, the United States, India, Russia, Japan, Germany, Korea, Canada, Iran and Saudi Arabia, and they represent 75% of global CO₂ emissions.

The aim of this article is to analyse the level of differentiation of the EU countries in terms of emissions of four types of greenhouse gases and to classify them into homogeneous groups. The article presents the taxonomic methods that allow the classification of various objects characterized by certain features. The subjects of this study are EU countries, and the features describing them are the emissions of GHG, such as: carbon dioxide, methane, nitrogen oxides, as the main "Kyoto gases" and additionally nitrous oxides, which are also an important family of air polluting chemical compounds. In this study one taxonomic method was used, this being cluster analysis. The groups of homogenous countries were distinguished by their total level of emissions and due to their level of emissions per capita. Analyses using agglomeration algorithms were conducted based on the annual Eurostat reports concerning the levels of GHG emissions of the EU member states. The resulting division of countries into clusters allows the criteria to be specified, directing special attention to the groups of countries constituting the greatest threat to the environment. Also, the result of grouping can act as the basis for a diversified policy of reducing GHG emissions in relation to particular clusters of EU countries. A further step may be to look for characteristic features of the countries included in the group.

Cluster analysis is used in many fields, e.g. in medicine to segment diseases or patients (Clinton, Button, Norring, & Palmer, 2004), in marketing for customer segmentation (Punj & Stewart, 1983; Wagner, Scholz, & Decker, 2005), and in the classification of countries or regions, according to various criteria (Bluszcz, 2016). It also applies in anthropology, archeology, and ethnography. Recently it has been widely used in research on artificial intelligence and pattern recognition. From the standpoint of GHG emissions, to our knowledge, there has only been one study which uses a non-hierarchical method, i.e. k-means (Kijewska & Bluszcz, 2016). In this study, we offer a hierarchical method – the method of agglomeration.

The rest of this article is organized as follows: discussion of cluster methods, in particular the method used for analysis – an agglomeration method, then the results of calculations, discussion and, finally, conclusions.

2. Methods

Researchers, scientists, and analysts are often faced with the need to organize a data set or various objects in a meaningful structure. Taxonomic methods in these situations are widely used because they only relate to the problems of classification and data analysis. The scope of this issue includes, among others: grouping objects, properties of taxonomic procedures, taxonomy structures, taxonomy time series, the theory of multivariate distributions and their mixtures, division optimization, classification and regression, graph theory and data mining. The following are among taxonomic methods (Figura, 2013):

- methods of structural classification; cluster analysis,
- methods of classification ranking – linear ordering of objects,
- methods of the identification of representatives of classes.

This article uses a structural classification method called cluster analysis, which uses methods and techniques that enable the extraction of objects from the data set that are similar to each other and to merge them into groups. As a result, from one heterogeneous set of objects a few homogeneous subsets can be obtained. These objects which are in the same set are considered "similar to each other" and objects from different subsets are treated as "dissimilar" (Mikut, 2009).

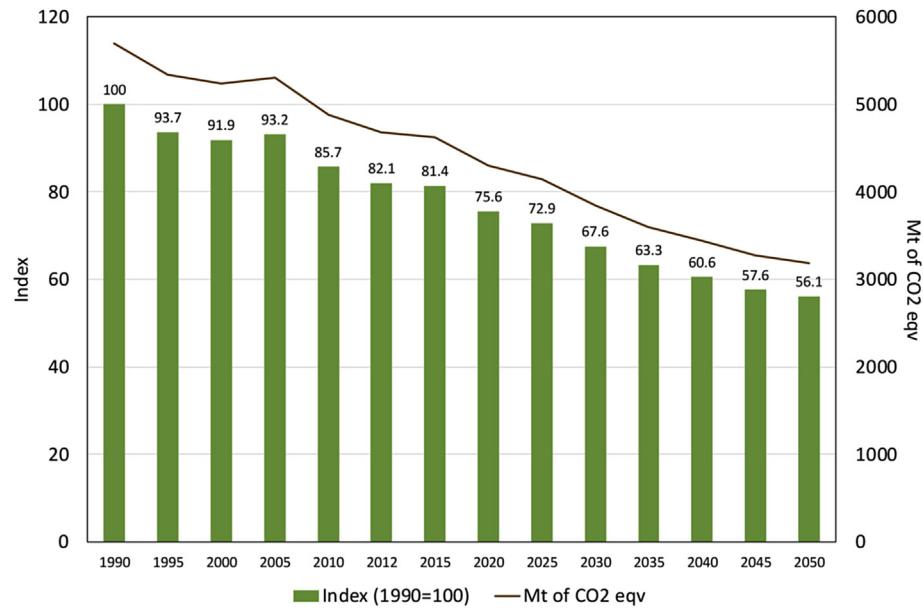


Fig. 4. Realized and projected GHG emissions in the EU for the years 1990–2050 (EC, 2013).

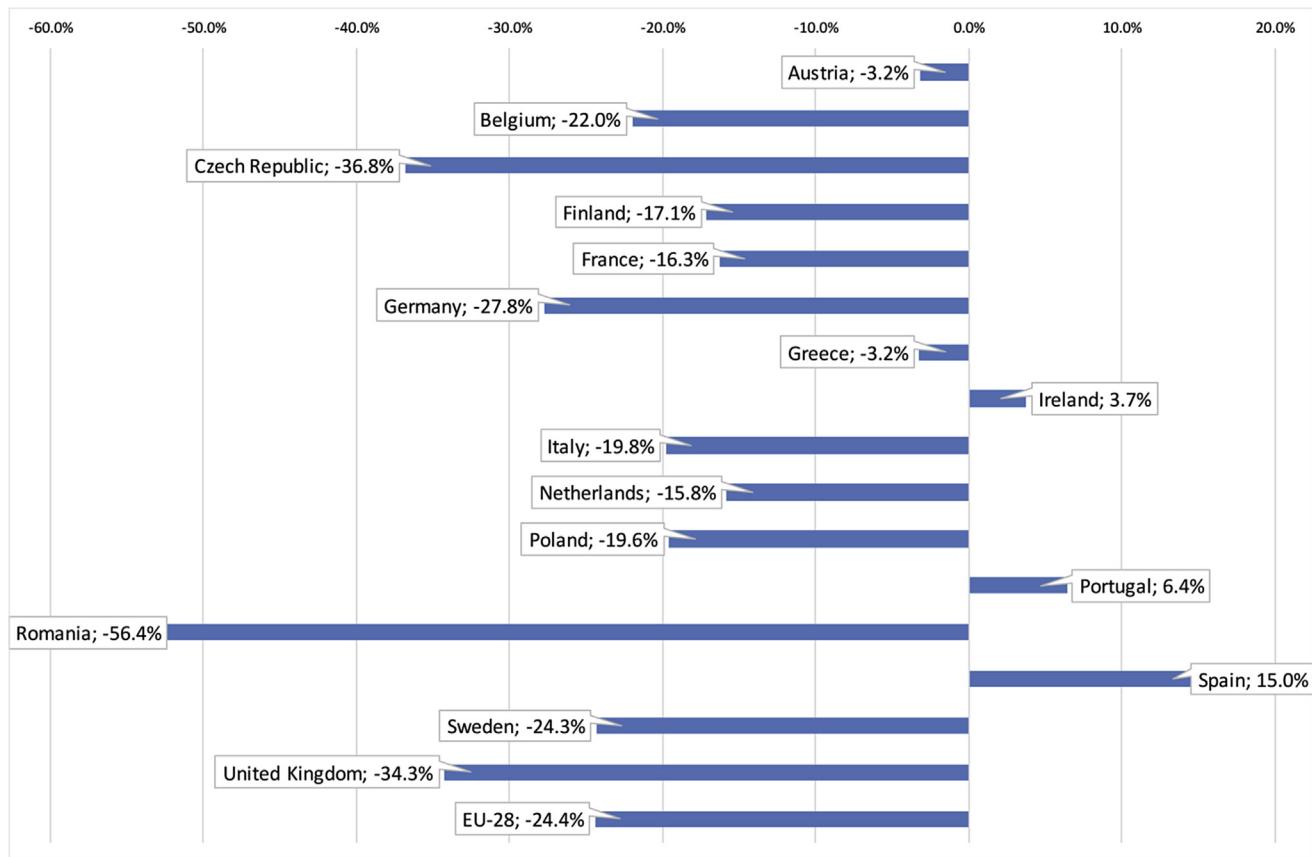


Fig. 5. Changes of GHG (CO₂ eq.) emissions in 16 EU countries from 1990 to 2014 (EEA, 2016).

The similarity criterion is either (a) distance: two or more objects belong to the same cluster if they are close according to a given distance – distance-based clustering – or (b) common concept: two or more objects belong to the same cluster if this cluster defines a concept common to all those objects – conceptual clustering. In clustering techniques, no *a priori* information about

classes is required; that is, neither the number of clusters nor the rules of assignment into clusters are known. They should be discovered exclusively from the given dataset without any reference to a training set (Girão, Postolache, & Pereira, 2009).

If we take distance as a criterion of similarity, then we can use different metrics. When all the components of the data set are in

the same physical units, the popular and relatively simple Euclidean distance can be used.

The Euclidean distance is calculated as the square root of the sum of the squared differences between the coordinates of a pair of objects (Giudici, 2003):

$$D_{XY} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (1)$$

where: $X = (x_{i1}, x_{i2}, \dots, x_{im})$, $Y = (x_{j1}, x_{j2}, \dots, x_{jm})$ are two m -dimensional objects.

Another version of the Euclidean distance is Squared Euclidean distance. It gives greater weight to objects that are further apart.

Among other popular distance metrics are: Manhattan Distance, Chebyshev Distance, and Minkowski Distance (Rao, Miller & Rao, 2008).

Minkowski distance is the generalized metric distance:

$$D_{XY} = \left(\sum_{k=1}^d |x_{ik} - x_{jk}|^p \right)^{1/p} \quad (2)$$

because for $p = 2$ the distance becomes the Euclidean distance, and for $p = 1$ it becomes city block distance. Chebyshev distance is a special case of Minkowski distance with $p = \infty$.

Cluster analysis involves many different algorithms. These can be broadly divided into hierarchical and non-hierarchical algorithms.

In hierarchical clustering, a treelike structure (dendrogram) is created through recursive partitioning (divisive method) or the combining (agglomerative method) of existing clusters. In the agglomerative clustering methods, initially it is assumed that each observation is a small group consisting of only a single element. Then, in succeeding steps, the two closest clusters are aggregated into a new combined cluster. In this way, the number of clusters in the data set is reduced by one at each step. Finally, all elements are combined into a single big cluster. Divisive clustering methods begin with all elements in one big cluster, with the most dissimilar elements being split off recursively, into a separate cluster, until

each element represents its own cluster. Among non-hierarchical k-means methods, probabilistic clustering and the methods of self-organizing are distinguished (Larose, 2005).

As previously mentioned, in the context of hierarchical cluster analysis, methods can distinguish agglomerative techniques and divisive techniques. In practice, the most commonly used are agglomerative methods. In these methods, we assume that a set of N objects that should be grouped and a distance matrix (or similarities) $N \times N$ are given. The algorithm can be expressed in several of the following steps (Giudici, 2003; Stanisz, 2007):

1. Assign each object to one cluster (N objects, N clusters).
2. Calculate the distances (similarities) between all objects. This can be done using different methods of linkage (ex. single linkage, complete linkage, the average linkage, weighted average linkage, the centroid method, or Ward's method).
3. Find the closest (most similar) pair of clusters and combine them into one single cluster; therefore, we get one cluster less.
4. Calculate the distances (similarities) between the new cluster and others.
5. Repeat steps 2, 3 and 4 until the moment when all objects are grouped together in one single cluster of size N ($n-1$ iterations).

As mentioned in step 2, clustering methods differ only in the way of calculating the distance between clusters. Below is a brief description of the most commonly used (Giudici, 2003; Mooi & Sarstedt, 2011). Graphic interpretation of some of the methods is shown in Fig. 6.

In the method, a single linkage or a nearest neighbourhood distance between two clusters is equal to the shortest distance between each object belonging to one cluster and each object belonging to another cluster. Thus, the distance between the two groups is defined as the minimum from distances between each of the n_1 observations belonging to the cluster C_1 and each of the n_2 observations belonging to the cluster C_2 :

$$d(C_1, C_2) = \min(d_{rs}) \text{ for } r \in C_1, s \in C_2. \quad (3)$$

In the method of complete linkage or farthest neighbourhood the distance between two clusters is equal to the largest distance

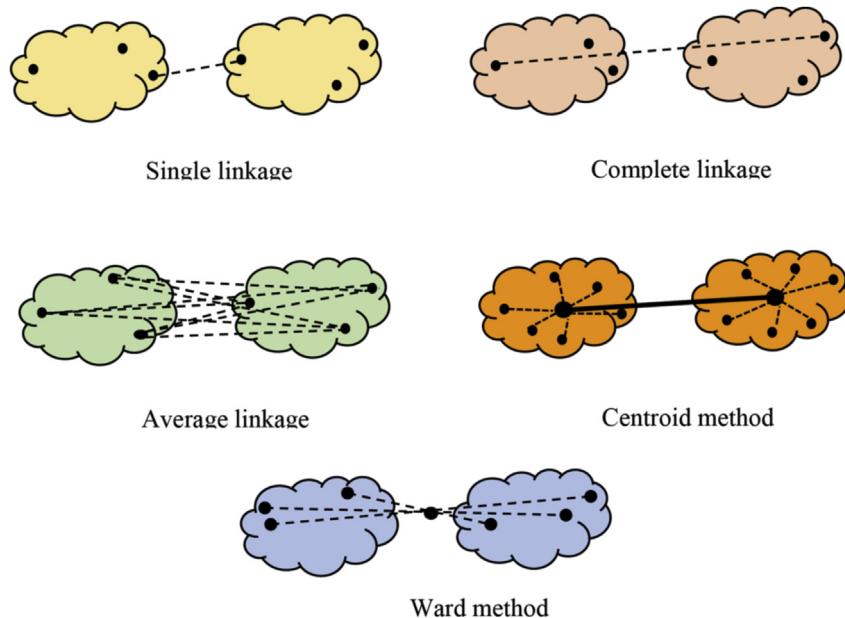


Fig. 6. Illustration of agglomerative methods of clustering (on the basis: Tan, Steinbach, & Kumar, 2005).

between each object belonging to one cluster and each object belonging to another cluster. Thus, the distance between the two groups is defined as the maximum distance between each n_1 observation belonging to cluster C_1 and each n_2 observation belonging to cluster C_2 :

$$d(C_1, C_2) = \max(d_{rs}) \text{ for } r \in C_1, s \in C_2. \quad (4)$$

In the average linkage method (unweighted pair-group average), it is assumed that the distance between one cluster and the second one is equal to the average distance from each element of one cluster to any element of the second cluster:

$$d(C_1, C_2) = \frac{1}{n_1 n_2} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} d_{rs} \text{ for } r \in C_1, s \in C_2. \quad (5)$$

The method of weighted pair-group average is similar to the previous one, except that in the calculation, the size of the cluster (i.e. the number of elements contained in it) is treated as the weight.

In the centroid method (unweighted pair-group centroid) the distance between two groups C_1 and C_2 , at respectively n_1 and n_2 observations, is defined as the distance between the respective centroids (usually the averages), \bar{x}_1 and \bar{x}_2 :

$$d(C_1, C_2) = d(\bar{x}_1, \bar{x}_2) \quad (6)$$

A centroid of a new cluster is calculated according to this formula:

$$\frac{\bar{x}_1 n_1 + \bar{x}_2 n_2}{n_1 + n_2} \quad (7)$$

It is worth noting the similarity between the average linkage method and the centroid method: the average linkage method considers the average distance between the observations of each of the two groups, whereas in the centroid method the centroid of each group is calculated and then the distance between the two centroids is calculated.

The method of weighted pair-group centroid (median) is identical to the centroid method, except that weighting is introduced into the calculations to consider differences in cluster sizes (i.e., the number of objects contained in them).

Ward's method minimizes the objective function using the principle that the aim is to create groups that have maximum internal cohesion and maximum external separation. This method attempts to minimize the error of the sum of squares (ESS) of any two (hypothetical) clusters, which can be created at each step. This method is generally considered to be very effective, however, it tends to form clusters of small size.

$$ESS = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \quad (8)$$

where x_{ij} is the j -th object in the i -th cluster, n_i is the number of objects in the i -th cluster.

Each of the algorithms described can lead to very different results based on the same set of observations. In general, it is believed that a single linkage algorithm is the most versatile. The complete linkage method is strongly influenced by outliers, as it is based on the maximum distances. Clusters built by this method can be quite compact and tightly clustered. The average linkage and centroid algorithms tend to form clusters with a relatively low within-cluster variance and similar sizes. However, both these methods are under the influence of outer elements, although this influence is not as strong as in the complete linkage method.

Ward's method is the most different among the aforementioned methods. When enlarging one of the clusters, the intragroup variance grows (calculated by the squares of deviations from the average in clusters); this method involves combining groups that provide the smallest increment of the variance for a given iteration. This method works well when clusters of relatively similar sizes are expected, and the set of observations does not contain outliers (Mooi & Sarstedt, 2011).

Since the presented algorithms are quite laborious, there are some programs such as STATISTICA, SPSS, STATA, ClustanGraphics, or others, which can be quite useful. By using them, cluster analysis, including the agglomerative method, can be performed. In these programs the results of clustering are illustrated in the form of dendograms.

The agglomerative method does not directly indicate what the optimal number of clusters is. Unfortunately, there is no objective rules to determine this number. Intuition, experience and substantive knowledge of the tested objects does matter, while some methods are proposed for estimating the number of clusters.

The easiest way is to analyse the obtained dendograms from the point of view of the differences between successive nodes. A large value of differences is tantamount to large cluster distances – in this point the division can be made. The graph of the course of the agglomeration can be used and this shows the distance between the clusters at the point of joining. A high jump on this chart is synonymous with great distance between each cluster. This jump is a place of division. There are also some formulas that were proposed by Grabiński or Mojena (Stanisz, 2007).

The research was performed to include four types of emission of greenhouse gases: carbon dioxide (CO_2), nitrogen oxides (NO_x), methane (CH_4) and nitrous oxide (N_2O). Using a STATISTICA package two perspectives were taken into consideration: the size of the total greenhouse gas emissions in the EU and the emissions in these countries per capita. Data was obtained from the website of EUROSTAT (2016). This data covers 25 EU countries (without Malta, Cyprus and Luxembourg), and additionally Norway, Switzerland and Turkey.

Each cluster method has its advantages and disadvantages. In this study, a hierarchical agglomerative method was selected. The choice was guided by several factors. Firstly, assuming in advance the number of clusters was not needed, as it is e.g. in the method of k-means. Secondly, the ability to visualize the results of grouping in the form of a dendrogram, which is more intuitively understandable. Moreover, if necessary, a different number of clusters can be chosen without the need for recalculations and a new interpretation of the results can be made. Of the five approaches to solving agglomeration algorithm (discussed previously), considering their advantages and limitations the complete linkage method was chosen. Using this method, the segmentation of EU countries into clusters will be conducted in terms of total emissions and emissions per capita.

3. Results and discussion

To achieve comparability between the data collected, normalization was conducted by standardization according to this formula (Morzy, 2013):

$$z_i = \frac{x_i - \bar{x}}{S_x} \quad (9)$$

where: \bar{x} i S_x are respectively, the mean and standard deviation of the variable in the sample.

Standardized values of greenhouse gas emissions by EU country for the year 2012 are presented in Table 1.

Table 1

Standardized data – greenhouse gas emissions by EU countries 2012.

	Carbon dioxide	Nitrogen oxide	Methane	Nitrous oxide
Austria	-0.45074	-0.55737	-0.61858	-0.48034
Belgium	-0.26866	-0.48805	-0.55921	-0.35700
Bulgaria	-0.53535	-0.64333	-0.51564	-0.48371
Croatia	-0.67841	-0.76728	-0.72129	-0.59808
Czech Republic	-0.28830	-0.47465	-0.36134	-0.31516
Denmark	-0.38828	1.22141	-0.60712	-0.37540
Estonia	-0.68845	-0.80347	-0.85780	-0.74664
Finland	-0.50811	-0.51778	-0.68558	-0.48752
France	1.00626	1.16150	1.85063	2.92740
Germany	3.62652	3.05892	1.75972	2.87713
Greece	-0.33457	-0.38551	-0.37673	-0.38012
Hungary	-0.53404	-0.60516	-0.47108	-0.36759
Ireland	-0.58632	-0.73703	-0.24819	-0.32733
Italy	1.18237	1.12854	0.99339	1.00767
Latvia	-0.73109	-0.79542	-0.81946	-0.69318
Lithuania	-0.68467	-0.72288	-0.74123	-0.53913
Netherlands	0.22991	0.05697	-0.08405	-0.31040
Norway	-0.50150	-0.36304	-0.67670	-0.59739
Poland	0.84919	0.80427	1.33690	1.12993
Portugal	-0.51838	-0.52170	-0.23468	-0.52126
Romania	-0.34609	-0.44327	0.30565	-0.05321
Slovakia	-0.60025	-0.72501	-0.67999	-0.62979
Slovenia	-0.69275	-0.77250	-0.80655	-0.74008
Spain	0.61926	0.99713	0.85954	0.38960
Sweden	-0.50929	-0.38697	-0.64676	-0.39816
Switzerland	-0.54012	-0.70511	-0.70688	-0.61320
Turkey	0.99272	1.21939	2.46250	0.15499
United Kingdom	1.87913	1.76740	1.85053	1.52795

As previously mentioned, the grouping of objects (countries) was conducted using the agglomerative algorithm with complete linkage. Incidentally Ward's method gave very similar results.

The division into clusters by the complete linkage method is as follows (see Fig. 7):

Cluster I – United Kingdom, France;

Cluster II – Turkey, Poland, Italy, Spain;

Table 2

Average total emissions of greenhouse gases in clusters (complete linkage method).

	Carbon dioxide [t]	Nitrogen oxide [t]	Methane [t]	Nitrous oxide [t]
Cluster I	448 530 308	1 215 002	2401 798	149 863
Cluster II	340 943 399	993 760	2 020 989	73 112
Cluster III	890 328 607	2 040 910	2322 716	181 874
Cluster IV	58 834 035	206 399	330 357	16 555

Cluster III – Germany;

Cluster IV – Denmark, Romania, Netherlands, Switzerland, Slovakia, Lithuania, Croatia, Latvia, Slovenia, Estonia, Portugal, Ireland, Hungary, Bulgaria, Norway, Sweden, Finland, Austria, Greece, Czech Republic, Belgium.

For each cluster the mean values of emissions were calculated. In this way, the characteristics of each cluster can be determined.

As can be seen from Table 2, grouping countries with the method of complete linkage, cluster III – Germany – has the highest CO₂, NO_x and NO₂. Emissions of CH₄ are at a similar level as in clusters I and III. In second place in terms of greenhouse gases is cluster I, and in third place is cluster II. Countries from cluster IV have the lowest emissions of all gases.

Greenhouse gas emissions, namely carbon dioxide, nitrogen oxide, methane and nitrous oxide has also been analysed from the perspective of emissions per capita. Standardized data is presented in Table 3.

Based on analysis of GHG emissions per capita, a different structure of homogeneous countries has been created. Also, in this analysis complete linkage was chosen. In this case Ward's method provided almost the same clusters. The analysis of dendograms allowed four clusters to be distinguished.

Thus, in a complete linkage method the following clustering was established (see Fig. 8):

Cluster I – Denmark;

Cluster II – Ireland;

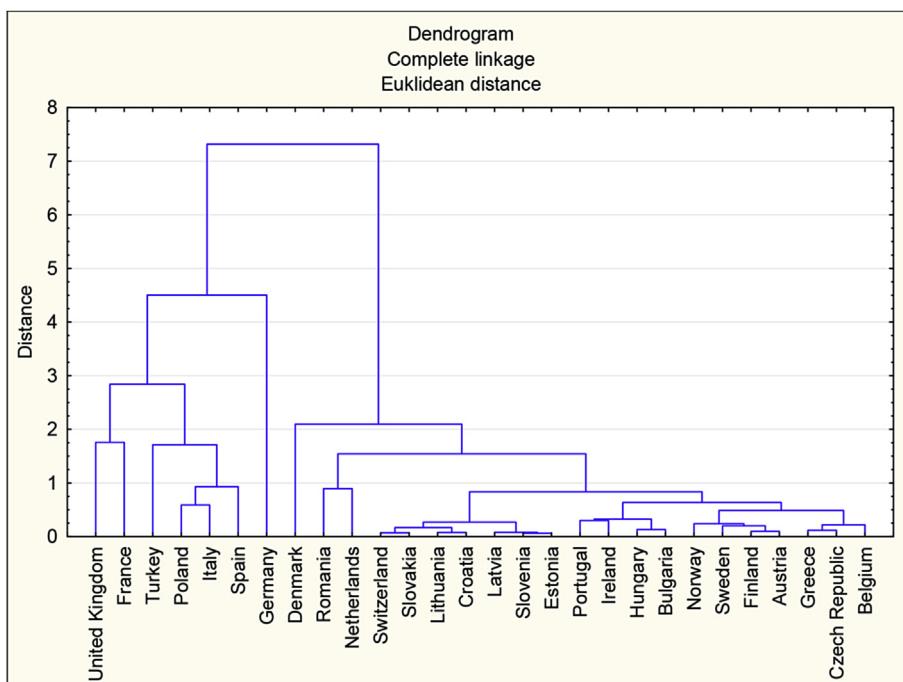


Fig. 7. Dendrogram of the grouping of EU countries according to total emissions by complete linkage.

Table 3

Standardized data - greenhouse gas emissions per capita in the EU in 2012.

	Carbon dioxide	Nitrogen oxides	Methane	Nitrous oxide
Austria	0.03013	-0.23262	-0.60179	-0.30889
Belgium	0.67064	-0.25579	-0.67776	-0.13673
Bulgaria	-0.49663	-0.34629	0.16056	-0.14235
Croatia	-1.16972	-0.42597	-0.25428	0.13957
Czech Republic	0.60281	-0.22591	0.22523	0.11494
Denmark	2.30094	4.92799	0.34845	1.79185
Estonia	1.92511	0.07326	-0.41006	0.24579
Finland	0.81578	0.20267	-0.27680	0.78479
France	-0.67795	-0.31964	-0.15778	0.74645
Germany	1.04136	-0.10211	-0.71705	-0.06343
Greece	0.07398	-0.14953	-0.01199	-0.36400
Hungary	-1.01486	-0.40071	-0.20131	-0.07972
Ireland	0.29216	-0.32472	4.63521	3.29690
Italy	-0.38891	-0.30809	-0.73284	-0.78875
Latvia	-1.37813	-0.25800	-0.42652	0.30380
Lithuania	-0.85021	-0.12148	0.04877	1.76561
Netherlands	1.55783	0.03021	0.08136	-0.80170
Norway	1.36211	0.81656	0.05347	-0.01265
Poland	0.28279	-0.15569	0.48879	0.25236
Portugal	-1.00669	-0.30776	0.70023	-0.95287
Romania	-1.30229	-0.50728	0.37703	-0.55122
Slovakia	-0.44105	-0.38128	-0.26521	-0.62532
Slovenia	0.15822	-0.00685	0.13187	-0.49915
Spain	-0.48639	-0.17993	-0.36810	-0.96497
Sweden	-0.68095	-0.01602	-0.89659	-0.05160
Switzerland	-0.51175	-0.46149	-0.96758	-0.99865
Turkey	-1.04805	-0.39475	-0.14742	-1.70184
United Kingdom	0.33972	-0.16879	-0.13788	-0.39823

Cluster III – Turkey, Romania, Portugal, Spain, Slovakia, Switzerland, Sweden, Italy, Lithuania, Latvia, Hungary, Croatia, France; Cluster IV – Netherland, Norway, Estonia, Finland, Poland, Czech Republic, Germany, Belgium, United Kingdom, Slovenia, Greece, Austria, Bulgaria.

As can be seen in Table 4, Denmark was divided into a separate cluster because it had the highest emissions per capita of CO₂ and

NO_x. Ireland, in turn, was spun off into a separate cluster because it had the highest emissions per capita of CH₄ – three times more than the average for all EU countries. Group III is characterized by the lowest values of all greenhouse gas emissions per capita

Although cluster analysis is widely used, it is difficult to find publications on the use of these methods for grouping countries in terms of greenhouse gas emissions. Kijewska and Bluszcz (2016) used the method of k-means for clustering EU countries with respect to GHG emissions. They also assumed the number of clusters to be equal to 4, however, the assignment of countries to groups is a little different from that shown in this study. On the other hand, Kolasa-Więcek (2013) uses the k-means method to group OECD countries in terms of agricultural GHG emissions. Fuzzy cluster analysis is used in (Xia et al., 2011) to group industrial sectors (in China), in terms of several indicators that describe energy security, efficiency and carbon emission. If we assume that the grouping of countries according to the criterion of GHG emissions is important then we do not ask whether cluster analysis should be used, but which method is best.

4. Conclusions

The agglomerative methods presented in this article are universal and can be widely used in business practice. The use of clustering takes place due to the following premises:

- reducing the large amount of collected information to a few basic categories, which enables easy orientation in a multidimensional phenomenon and the establishment of a typology in terms of the study issues and reasoning;
- the identification of homogeneous objects of analysis, in which it is easier to extract systematic factors and possible causal relationships;
- reducing the amount of time and research costs by limiting consideration to the most important facts with little loss of information and the relatively low probability of receiving the results of distorted analysis.

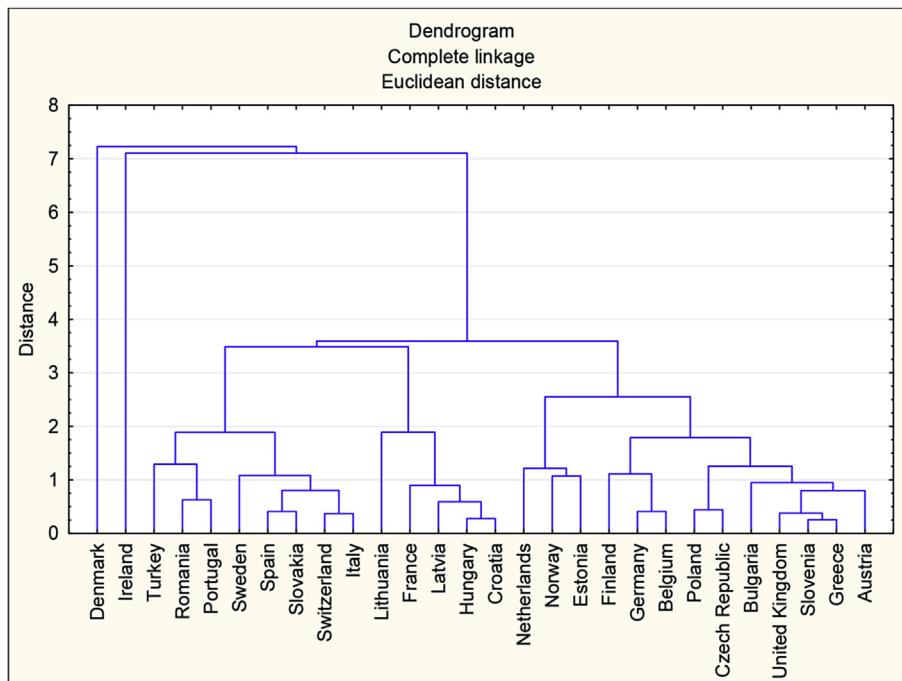


Fig. 8. Dendrogram of clustering EU countries according to emissions per capita by complete linkage.

Table 4

Average greenhouse gas emissions per capita in clusters by complete linkage.

	Carbon dioxide [t/capita]	Nitrogen oxides [t/capita]	Methane [t/capita]	Nitrous oxide [t/capita]
Cluster I (Denmark)	14 466 164	201 687	48 618	3992
Cluster II (Ireland)	8 648 212	16 970	130 672	5438
Cluster III	5 360 965	17 346	37 101	1993
Cluster IV	9 665 417	26 980	39 456	2173

The results of the analysis complete information about the different levels of greenhouse gas emissions in the European Union. The classification of EU countries in terms of total greenhouse gas enables the determination of groups of countries with similar emissions, the determination of the largest issuers and the group of countries which contribute least to pollution. On the other hand, information about GHG emissions per capita is equally important. As demonstrated, the resulting clustering do not coincide with those concerning total emissions of GHG. Among the largest emitters of greenhouse gases are Germany, the United Kingdom, France, Turkey, Poland, Italy and Spain. However, in the classification of gas emitters per capita the negative leaders are Denmark and Ireland. The latter has been classified as a single cluster mainly due to its high methane emissions.

Poland is in a group of average issuers. It is worth noting that considering emissions per capita of CO₂, NO_x and NO₂ Poland has a lower value than for example Germany, but in the whole cluster No. IV has the highest methane emissions per capita.

The resulting division of countries into homogeneous clusters enables the specification of the criteria that influenced such a division, special attention to be directed towards the group of countries which constitute the greatest threat to the environment. This division of countries should also be the basis for further debate in the context of the methodology of strategic guidelines for granting emission limits to EU countries.

References

- Bluszcz, A. (2016). Classification of the European Union member states according to the relative level of sustainable development. *Quality & Quantity*, 50(6), 2591–2605. <http://dx.doi.org/10.1007/s11135-015-0285-y>.
- Bluszcz, A., & Kijewska, A. (2015). Challenges of sustainable development in the mining and metallurgy sector in Poland. *Metallurgy*, 54(2), 441–444.
- CDIAC. (2016). Carbon dioxide information analysis center, Oak ridge national laboratory. Oak Ridge, Tenn., United States: US Department of Energy. Retrieved September 20, 2016 from http://cdiac.ornl.gov/ftp/ndp030/global.1751_2013.ems.
- Ciepielka, D. (2011). Rynek, energia i klimat [Market energy and climate]. Nowy Przemysł, (5). Retrieved May 15, 2015 from http://www.wnp.pl/wiadomosci/rynek-energia-i-klimat,-7073_1_0_0_2.html.
- Climate Policy of Poland. (2003). Strategies to reduce greenhouse gas emissions in Poland by 2020. Warsaw: The Ministry of the Environment. Retrieved March 20, 2015 from www.chronmyklimat.pl/download.php?id=107.
- Clinton, D., Button, E., Norring, C., & Palmer, R. (2004). Cluster analysis of key diagnostic variables from two independent samples of eating-disorder patients: Evidence for a consistent pattern. *Psychological Medicine*, 34, 1035–1045. <http://dx.doi.org/10.1017/S0033291703001508>.
- Dubiński, J., & Turek, M. (2014). Chances and threats of hard coal mining development in Poland – the results of experts research. *Archives of Mining Science*, 59(2), 395–411. <http://dx.doi.org/10.2478/amsc-2014-0028>.
- EC (2013). European Commission. EU Energy, Transport and GHG Emissions. Trends to 2050. Reference scenario 2013. Publication Office of the European Commission 2013. Retrieved March 20, 2015 from http://ec.europa.eu/clima/policies/2030/docs/eu_trends_2050_en.pdf.
- EEA (2016). European Environment Agency, Retrieved January 27, 2017 from <http://www.eea.europa.eu/data-and-maps/data/data-viewers/greenhouse-gases-viewer>.
- Eurostat (2016) The source data for GHG emissions. Retrieved October 10, 2016 from <http://ec.europa.eu/eurostat/data/database>.
- Figura, J. (2013). *Taksonomia w polityce logistycznej państwa* [Taxonomy in logistic State policy]. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
- Girão, P. S., Postolache, O., & Pereira, J. M. D. (2009). Data fusion, decision-making, and risk analysis: Mathematical tools and techniques. In F. Pavese, & A. B. Forbes (Eds.), *Data, modeling for metrology and testing in measurement science* (pp. 1–50). Boston: Birkhäuser.
- Giudici, P. (2003). *Applied data mining. Statistical methods for business and industry*. West Sussex: John Wiley & Sons.
- Idso, C. D., Carter, R. M., Singer, S. F., & Soon, W. (2013). *Scientific critique of IPCC's summary for policymakers*. Center for the study of carbon dioxide and global change. The Heartland Institute, Science and Environmental Policy Project. Retrieved March 20, 2015 from https://www.heartland.org/sites/default/files/critique_of_ipcc_spm.pdf.
- IEA (2016). International Energy Agency. CO₂ Emissions from Fuel Combustion 2016 Edition. Retrieved September 20, 2016 from https://www.iea.org/bookshop/729-CO2_Emissions_from_Fuel_Combustion.
- Jonek-Kowalska, I. (2014). Risk management in the hard coal mining industry: Social and environmental aspects of colliers' liquidation. *Resources Policy*, 41, 124–134. <http://dx.doi.org/10.1016/j.respol.2014.05.002>.
- Jonek-Kowalska, I. (2015). Challenges for long-term industry restructuring in the Upper Silesian Basin. What has polish coal mining achieved and failed from a twenty-year perspective? *Resources Policy*, 44, 135–149. <http://dx.doi.org/10.1016/j.respol.2015.02.009>.
- Karbownik, A., & Stachowicz, J. (1994). Social aspects of restructuring hard coal mining in Poland. *Resources Policy*, 20(3), 198–201. [http://dx.doi.org/10.1016/0301-4207\(94\)90052-3](http://dx.doi.org/10.1016/0301-4207(94)90052-3).
- Karbownik, A., Turek, M., & Pawełczyk, E. (2001). Efekty realizacji rządowego programu reformy górnictwa węgla kamiennego w latach 1998–2001 [The effects of implementation of the government program of reform of coal mining in 1998–2001]. In J. Pyka (Ed.), *Globalne i regionalne uwarunkowania nowoczesności przemysłu i usług* [Global and regional conditions of modern industry and services]. Katowice: Wyd. Akademii Ekonomicznej im. Karola Adamieckiego.
- Kijewska, A., & Bluszcz, A. (2016). Research of varying levels of greenhouse gas emissions in European countries using the k-means method. *Atmospheric Pollution Research*, 7(5), 935–944. <http://dx.doi.org/10.1016/j.apr.2016.05.010>.
- Kolasa-Więcek, A. (2013). The use of cluster analysis in the classification of similarities in variables associated with agricultural greenhouse gases emissions in OECD countries. *Wieś i Rolnictwo*, 1(158), 59–66.
- Korban, Z., & Manowska, A. (2011). Wykorzystanie ciągów czasowych w procesie szacowania poziomu emisji dwutlenku węgla [The use of timesequences in the process of estimating carbon dioxide emissions]. *Mining and Geology*, 6(4), 39–48.
- Korski, J., Tobór-Osadnik, K., & Wyganowska, M. (2015). Perfect model of mining equipment user requirement by C. Conley. *Management Systems in Production Engineering*, 18(2), 94–97. <http://dx.doi.org/10.1016/j.respol.2016.02.005>.
- Korski, J., Tobór-Osadnik, K., & Wyganowska, M. (2016). Reasons of problems of the Polish hard coal mining in connection with restructuring changes in the period 1988–2014. *Resources Policy*, 48, 25–31.
- Krause, E., Krzemień, A., & Smoliński, A. (2015). Analysis of assessment of a critical event during an underground coal gasification experiment. *Journal of Loss Prevention in the Process Industries*, 33, 173–182. <http://dx.doi.org/10.1016/j.jlp.2014.12.014>.
- Krzemień, A., Więckol-Ryk, A., Duda, A., & Kotera, A. (2013). Risk assessment of a post – combustion and amine – based CO₂ capture ready process. *Journal of Sustainable Mining*, 12(4), 18–23. <http://dx.doi.org/10.7424/jsm130404>.
- Laing, T., Sato, M., Grubb, M., & Comberti, C. (2013). Assessing the effectiveness of the EU Emissions Trading System. Centre for Climate Change Economics and Policy Working Paper No. 126 Grantham Research Institute on Climate Change and the Environment Working Paper No. 106. Retrieved January 27, 2017 from <http://www.lse.ac.uk/GranthamInstitute/wp-content/uploads/2014/02/WP106-effectiveness-eu-emissions-trading-system.pdf>.
- Larose, D. T. (2005). *Discovering knowledge in data. Introduction to data mining*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Lorek, E. (2011). Efektywność europejskiego i międzynarodowego systemu handlu emisjami jako instrumentu zrównoważonej gospodarki energetycznej [The effectiveness of European and international emissions trading system as an instrument for sustainable energy]. *Przegląd Górnictwa*, 67(9), 103–106.
- Lutyński, M. (2014). Impact of preparation and storage of activated carbon on the high pressure sorption of CO₂. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 62(1), 113–119, 10.2478/bpasts-2014-0013.
- Lutyński, M., Sakiewicz, P., & Gonzalez, M. A. (2014). Halloysite as mineral adsorbent of CO₂ – kinetics and adsorption capacity. *Journal of the Polish Mineral Engineering Society*, 15(1), 111–117.

- Mikut, G. (2009). Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku [Application of cluster analysis and decision trees for segmentation of the market]. *StatSoft Poland*, 75–92. Retrieved March 20, 2015 from www.statsoft.pl/czytelnia.
- Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research*. Berlin-Heidelberg: Springer-Verlag.
- Morzy, T. (2013). *Eksploracja danych. Metody i algorytmy [Exploration of data. Methods and Algorithms]*. Warszawa: PWN.
- Pachauri, R. K., & Meyer, L. A. (2014). *Climate Change 2014. Synthesis Report*. Retrieved March 20, 2015 from http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full.pdf.
- Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2), 134–148. <http://dx.doi.org/10.2307/3151680>.
- Ranoss, R. (2008). Organization and trade on CO₂ emission allowances market. *Polityka Energetyczna – Energy Policy Journal*, 11(2), 85–95.
- Rao, C. R., Miller, J. P., & Rao, D. C. (Eds.). (2008). *Handbook of Statistics: Epidemiology and medical statistics*. North-holland. Elsevier.
- Stanisz, A. (2007). *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. T. 3 Analizy wielowymiarowe [Accessible statistics course using STATISTICA PL with examples from medicine. Vol. 3. Multivariate Analyses]*. Krakow: StatSoft.
- Stańczyk, K., Dubiński, J., Cybulski, K., Wiatowski, M., Świądrowski, J., Kapusta, K., et al. (2010b). Underground coal gasification – international experiments and experiments in Barbara mine. *Energy Policy*, 13(2), 423–433.
- Stańczyk, K., Kapusta, K., Wiatowski, M., Świądrowski, J., Smoliński, A., & Rogut, J. (2012). Experimental simulation of hard coal underground gasification for hydrogen production. *Fuel*, 91(1), 40–50. <http://dx.doi.org/10.1016/j.fuel.2011.08.024>.
- Stańczyk, K., Smoliński, A., Kapusta, K., Wiatowski, M., Świądrowski, J., Kotyrba, A., et al. (2010a). Dynamic experimental simulation of hydrogen oriented underground gasification of lignite. *Fuel*, 89(11), 3307–3314. <http://dx.doi.org/10.1016/j.fuel.2010.03.004>.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to data mining*. Boston: Addison-Wesley.
- UNFCCC. (2008). *Kyoto Protocol reference manual. On accounting of emissions and assigned amount*. United Nations Framework Convention on Climate Change. Retrieved January 27, 2017 from http://unfccc.int/kyoto_protocol/items/3145.php.
- Wagner, R., Scholz, S. W., & Decker, R. (2005). The number of clusters in market segmentation. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data analysis and decision support* (pp. 157–176). Heidelberg: Springer.
- Wiatowski, M., Stańczyk, K., Świądrowski, J., Kapusta, K., Cybulski, K., Krause, E., et al. (2012). Semi-technical underground coal gasification (UCG) using the shaft method in experimental Barbara mine. *Fuel*, 99, 170–179. <http://dx.doi.org/10.1016/j.fuel.2012.04.017>.
- Xia, X. H., Huang, G. T., Chen, G. Q., Zhang, Bo, Chen, Z. M., & Yang, Q. (2011). Energy security, efficiency and carbon emission of Chinese industry. *Energy Policy*, 39, 3520–3528. <http://dx.doi.org/10.1016/j.enpol.2011.03.051>.