CLASSIFICATION OF EU COUNTRIES IN TERMS OF THE EVOLUTION OF THE GHG INDICATOR USING CLUSTER ANALYSIS

MARINOIU CRISTIAN
ASSOC. PROF.PH.D., PETROLEUM-GAS UNIVERSITY OF PLOIEŞTI
e-mail:marinoiu_c@yahoo.com

Abstract

Greenhouse gases are one of the main factors that influence the Earth's global temperature variation. As the result of both the beginning of the industrial revolution (the 1750's) and the intensificication and diversification of human activities, the volume of greenhouse gasses increases significantly. The risk of an accelerated global warming can be decreased by reducing the volume of greenhouse gasses emissions resulting from human activities. The annual volume of these emissions is reflected by the Greenhouse gas (GHG) indicator. This work carries out a classification of EU countries on the basis of the evolution of the GHG indicator using Partitioning Around Medoids (PAM) method.

Key words: *greenhouse gas, cluster analysis, PAM method*

Classification JEL : *C38, Q54*

1. Introduction

Climate changes are caused mainly by the following factors [6]:

- Change in the volume of greenhouse gases emitted each year;
- Variation of solar radiation that reaches the Earth;
- The variance in the reflection power of both the surface and the atmosphere of the Earth.

GreenHouse Gas (GHG) are "gases that trap heat in the atmosphere" [7]. Greenhouse gases which influence the climate changes to a large extent are: carbon dioxide, methane, nitrous oxide. The variation in the volume of greenhouse gases emitted into the atmosphere each year can be explained by both natural causes and as a consequence of human activities. The influence of human activities started to be significant since the 1750's, with the onset of the industrial revolution, as shown by specialized studies [6]. At the same time the main human activities that have led to the accelerated increase in the volume of greenhouse gas emissions have been identified : exploitation, transport and combustion of fossil fuels (coal, oil and their derivatives) as well as extensive and intensive exploitation of the land. The obvious climate changes (the annual increase in the average temperature of the planet, the accelerated reduction of polar ice-boats, etc.) have alerted the international community, determining the responsible authorities to propose measures to stop this hapless evolution. One important aspect related to this is the Kyoto Protocol adopted in Kyoto in Japan, on 11 December 1997 by 37 industrialized countries and the European Community (15 Member countries) and which came into force on 16 February 2005. The Protocol stipulated, for the signatory nations annual emission limitations of greenhouse gases that would achieve the targets at the end of some agreed periods of time. It should be noted that the EU has fulfilled its commitments.

Taking all these considerations into account we believe that a classification of the EU countries from the perspective of the annual evolution of GHG indicator can offer a synthesized and useful view of the EU politics in this important field. In order to accomplish our purpose we use the cluster analysis. This powerfull instrument of classification was also used by other authors preoccupied by the identification of similarities and dissimilarities between countries regarding the politics which concern the greenhouse gas emissions.

A classification of the European Union countries that uses cluster analysis based on the GHG indicator is shown in paper [2]. The author presents his results in the context of comparative analysis of four environment indicators during 2002-2014: total greenhouse gas emissions, share of renewable energy in gross final energy consumption, primary energy consumption and final energy consumption. For the analysis, the author considers as representative the values of GHG for the following years: 2007, 2012 and 2014. A case study on the trend followed by 23 countries after the implementation of the protocol from Kyoto is presented in [1]. For this study, among other specific variables, the authors also consider GHG indicator. In order to discover the groups of countries with similar evolution during 2008-2012, the authors have used a cluster analysis technique called fuzzy C-medoids.

In the article [4] the author developes "a countries typology according to their GHG behavior and socio – economic variables causing these emissions", using principal component analysis and cluster analysis.

2. The GHG indicator

In this paper we use the data of GHG indicator. GHG indicator values used for the classification of the European Union countries are based on the Eurostat [8] The most relevant details of this indicator, that have to be taken into account, are as follows:

- The indicator relates to the annual volume of greenhouse gas emissions (which includes carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and the so-called F-gases (hydrofluorocarbons, perfluorocarbons and sulphur hexafluoride (SF6)), taking as a basis for reporting the year 1990;
- The indicator includes Greenhouse gas emissions from international aviation, but it does not include emissions and removals related to land use, land-use change and forestry (LULUCF) and emissions from international maritime transport;
- The time period that we refer to is 1991-2012.

Annual GHG indicator values associated with each of the EU countries make up a time series. In table 1 there are some examples of such a time series.

Table 1. Annual GHG indicator values for the 12 countries [7] of the European Union for the period 1991-2012

	BE	BG	CZ	DK	DE	EE	IE	EL	ES	FR	HR
1991	101.03	79.01	92.86	114.83	96.26	92.20	101.31	99.31	103.24	104.12	78.56
1992	100.16	73.71	84.47	106.33	92.38	67.34	101.05	100.74	106.41	102.71	73.46
1993	99.50	72.75	81.29	109.61	91.72	52.32	102.39	100.02	102.80	98.01	74.04
1994	103.39	68.99	76.25	115.50	90.20	53.89	104.66	103.00	108.56	98.18	71.55
1995	104.91	69.74	77.46	110.81	89.91	49.40	106.64	104.61	113.83	99.62	73.38
1996	107.94	69.34	79.30	129.24	91.48	51.01	110.19	107.35	111.48	102.32	75.36
1997	102.21	65.92	77.44	115.78	88.69	50.09	113.27	111.76	116.26	101.43	79.88
1998	106.32	61.41	73.86	110.56	86.69	46.31	118.32	117.08	119.97	104.25	80.18
1999	102.38	55.05	70.00	106.90	84.11	43.02	120.47	117.23	129.05	101.92	84.05
2000	103.05	54.36	74.71	100.72	84.12	42.29	124.35	120.21	134.84	101.57	83.00
2001	102.29	57.31	74.73	102.87	85.26	43.20	128.57	120.91	133.79	101.26	86.53
2002	101.49	54.64	72.92	101.66	83.57	41.72	125.50	120.85	139.84	100.33	89.85
2003	102.11	59.05	74.54	108.63	83.47	46.33	125.63	125.06	142.63	101.29	94.07
2004	102.79	58.31	75.39	100.64	82.62	47.19	124.91	125.55	148.09	101.19	94.73
2005	99.69	58.52	74.74	94.70	80.76	45.60	128.15	128.33	153.24	101.51	95.76
2006	97.25	59.22	75.29	105.89	81.49	44.04	127.96	125.41	150.81	99.58	97.53
2007	94.09	62.79	75.44	99.49	79.51	51.82	126.86	128.11	153.93	98.05	102.17
2008	95.94	61.43	72.90	94.55	79.79	48.21	125.85	124.61	142.33	97.22	98.10
2009	87.04	52.97	68.79	90.05	74.40	40.00	114.64	118.02	128.57	92.80	91.75
2010	92.26	55.33	70.18	90.67	77.06	49.13	114.04	111.73	124.41	94.08	90.26
2011	85.27	60.54	69.29	83.84	75.58	50.56	106.25	108.97	124.41	89.52	89.21
2012	82.56	56.02	67.32	79.63	76.55	47.40	107.04	105.71	122.48	89.46	82.65

Source: made by the author using data from Eurostat [8]

In this paper we used cluster analysis [5] to classify the time series and the EU countries in terms of the evolution of the GHG indicator.

Cluster analysis benefits from the existence of a multitude of methods that allow the discovery of the clusters in a set of data. Among the most well-known clustering methods are those based on partitioning and ranking. For cluster analysis we have chosen to use PAM algorithm [5].

3. PAM method for clustering

PAM (Partitioning Around Medoids) algorithm is, as suggested by its name part of the class of algorithms based on partitioning methods. The algorithm realises the partition of a set of objects C into disjoint classes called clusters. Each obtained cluster will contain objects which are similar from a certain point of view while the objects located in different clusters are disimilare [5]. PAM algorithm known as the k-medoids is inspired by the k-means algorithm, an algorithm for partitioning in which clusters are represented through their centroids [5]. In the PAM algorithm the clusters are represented by objects of the set C called medoids. The medoid is that object of a cluster for which the sum of distances of the other cluster objects to it, is minimal. This new approach makes the PAM algorithm an algorithm more robust to the outlieres than k-means algorithm and, unlike the latter, applicable both to data coming from the field of continuous and discrete domain.

Using PAM method involves knowing the number of clusters k and establishing a measure of dissimilarity between two objects.

4. Applying PAM method, results and discussions

Determining the optimal number of clusters k can be done through various techniques: the silhouette method, gap statistics method, the sum of squared error (SSE) scree plot method, the criterion of Calinsky etc.

In this paper we use the silhouette method [9]: the optimal structure of clusters is the structure for which the value of the average silhouette width is maximum. In table 2 these values are presented for different values of the number of clusters k. It can be observed that the maximum value of the of the average silhouette width, named silhouette coefficient, is 0.61 and this is obtained for k = 3.

Table 2. Average silhouette widths based on the number of clusters

Number of clusters	2	3	4	5	6	7	8	9	10
Average silhouette width	0.59	0.61	0.48	0.50	0.51	0.49	0.45	0.42	0.39

Source: made by the author with results obtained in R

The result obtained is reinforced by the fact that when applying the sum of squared error (SSE) scree plot method (figure 1) the same optimal value k = 3 for the number of clusters is suggested.

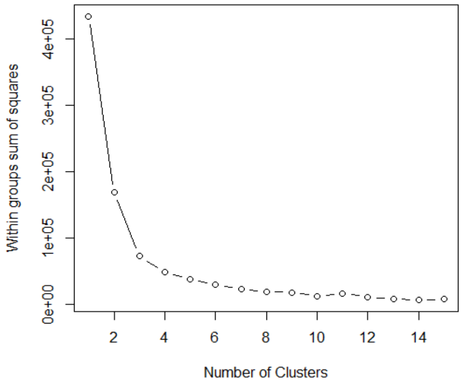

Figure 1. The sum of squared error (SSE) scree plot

Source: made by the author in R,using data from [8]

In the interpretation given in [9] the value 0.61 of the silhouette coefficient indicates that for k = 3 the structure of clusters found is reasonable. The appropriate silhouette of this choice is displayed in figure 2.

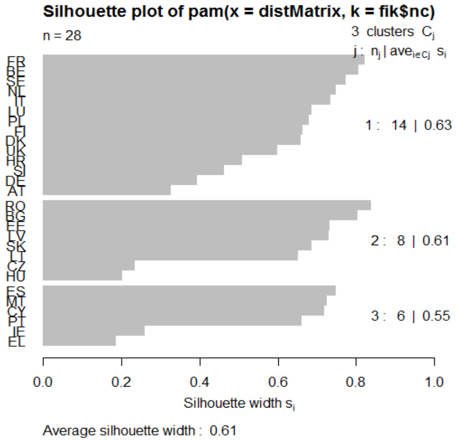

Figure 2. Silhouette plot for optimal number of clusters k=3

Source: made by the author in R

The GHG time series for each country is labeled by the international symbol of that country [10]. Although GHG time series can be designed as a point in the 22-dimensional space, the dynamic structure, as well as the autocorrelation of their values make that the use of the Euclidian distance (as a measure of the dissimilarity between them) should not be considered a good choice. The DTW distance [11], used in this paper as a measure of the dissimilarity between GHG time series, is one of the viable alternatives commonly used.

The graph of the silhouette in figure 2 highlights both the quality of the found structure and the composition of each cluster.

However, a much more suggestive graphic representation of the clusters is given in figure 3. Here, the elements of each cluster are plotted with respect to the first two principal components [3]. The fact that the two main components cover over 95% of the data variability reflects the very good quality of the 22-dimensional space approximation through the 2-dimensional space. Each cluster is bounded by an ellipse outline having a minimum area and, in order to highlight the distances between clusters, the centres of ellipses are joined by lines. The distance between the hatching lines reflects the density of objects in the cluster, and it is determined as the ratio between the number of objects in the cluster and the ellipse area which delimitates the cluster: the greater the density, the smaller the distance between the hatch lines.

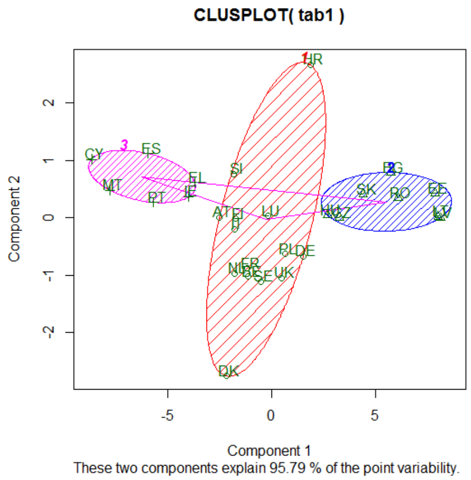

Figure 3. Obtained clusters versus first two principal components

Source: made by the author in R,using functions in R

The structure of clusters found is the following:

Cluster 1: France, Belgium, Sweden, the Netherlands, Italy, Luxembourg, Finland, Denmark, Poland, England, Croatia, Slovenia, Germany, Austria

Cluster 2: Romania, Bulgaria, Slovakia, Lithuania, Latvia, Estonia, the Czech Republic, Hungary

Cluster 3: Spain, Malta, Cyprus, Portugal, Ireland, Greece

In table 3 are presented the means and standard deviations of the GHG time series values for each cluster obtained.

Table 3. Means and standard deviations of the GHG time series values for each cluster obtained

	Mean	Standard deviation
Cluster1	97.29	10.24
Cluster2	61.63	14.50
Cluster3	127.33	16.96

Source: made by the author using results obtained in R

The table reveals a high mean of GHG indicator for countries that form the cluster 3. Basically this value is double, for the mean of the GHG indicator corresponding to the countries which form the cluster 2. This is reflected by the time-series graphs corresponding to countries which belong to the two clusters (figure 3).

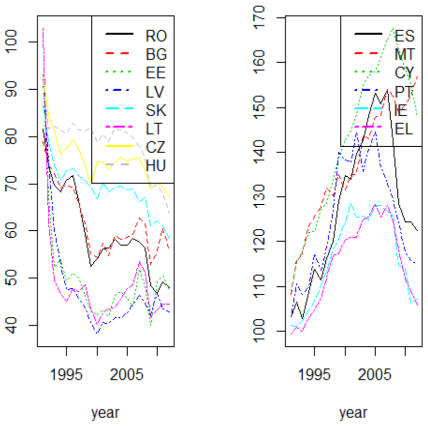

Figure 4. Left: GHG time-series graph for countries that belong to the cluster 2
Right: GHG time-series graph for countries that belong to the cluster 3
Source: made by the author in R, using functions in R

5. Conclusions

This paper carries out a classification of EU member countries according to the GHG indicator trends for each country over the period 1991 to 2012. For this classification it was used the PAM method, which is one of the most powerful techniques of cluster analysis. The performed cluster analysis highlights the existence of 3 well-defined clusters that refelect the discrepancies and similarities between the evolution of GHG indicator among the European Union countries. We can notice the similar evolution of this indicator in the case of former socialist countries (cluster 2). Notable exceptions are the cases of Poland, Slovenia and Croatia. These countries belong to cluster 1 which includes the economically well-developed countries of the European Union. The most significant discrepancy can be noticed between the groups of countries that belong to cluster 2 and those that belong to cluster 3: in cluster 2 the mean of GHG indicator is two times smaller than the mean associated with cluster 3.

6. Bibliography

[1] **Coppi R., De Lauretis R, D'Urso P., Romano D**.,*Statistical Analysis of Greenhouse Gas Emission Trends under the Kyoto Protocol,* In: S.I.S. 2005 - Statistica e Ambiente, 21-23 Settembre 2005, Messina, Italy , available at http://cab.unime.it/mus/377
[2] **Ionescu R.V**.,*Regional Environment Disparities and Europe 2020 Strategy's Goals,* Performance and Risks in the European Economy, EIRP Proceedings, Vol 9 (2014), available at http://www.proceedings.univ-danubius.ro/index.php/eirp/article/view/1507
[3] **Pisonn G., Struyf A.,. Rousseeuw P., J.**, *Displaying a clustering with CLUSPLOT*, Computational Statistics & Data Analysis, volume 30, issue 4, pp. 381-392, Elsevier,1999
[4] **Pita M. F**., *Towards a typology of countries according to their greenhouse gases emissions behavior,* Boletin de la Associacion de Geografos Espanoles, nr 59 (2012), p.459-463
[5] **Theodoridis S., Koutroumbas K**., Pattern Recognition, Academic Press Elsevier, 2009
[6] *** http://www.epa.gov/climatechange/science/causes.html
[7] *** http://www.epa.gov/climatechange/ghgemissions/gases.html
[8] *** http://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tsdcc100
[9] *** http://www.jstatsoft.org/v01/i04/paper
[10] *** http://publications.europa.eu/code/pdf/370000en.htm#pays
[11] *** http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_769.pdf