K-MEDOIDS AND FUZZY C-MEANS ALGORITHMS FOR CLUSTERING CO_2 EMISSIONS OF TURKEY AND OTHER OECD COUNTRIES

OZDEMIR, O.* – KAYA, A.

Department of Statistics, Faculty of Science, Anadolu University, 26470, Eskisehir, Turkey
(phone: +90-222-335-0580; fax: +90-222-320-4910)

**Corresponding author*
e-mail: ozerozdemir@anadolu.edu.tr

(Received 15th Jan 2018; accepted 16th Apr 2018)

Abstract. One of the most fundamental problems of societies is environmental problems that have emerged as the result of human behaviors which can be summarized as "dominating nature" and "using nature carelessly". The main causes of climate change and greenhouse gas emissions, which are considered to be most important environmental problem nowadays, are used of fossil fuel resources intensively in order to overcome the energy requirement. Carbon dioxide (CO_2) is the largest share of greenhouse gas emissions. To reduce CO_2 emissions, significant studies have been done in the world. In this study, we tried to classify member countries of OECD according to CO_2 emissions indicators from fossil fuel consumption. For the clustering, both the classical cluster analysis and the fuzzy cluster analysis were implemented. The clusters obtained from both methods are simply interpreted.
Keywords: *fuzzy clustering, Paris climate agreement, Kyoto Protocol, climate change, greenhouse gas emission*

Introduction

Climate change can be described in a very general way in the form of "regardless of their cause, long-term and slow-growing changes in climate conditions" (Türkeş, 2003).

However; with the industrial revolution, in addition to natural climate variability, it has entered a new era of human activities that affect the climate. Since the industrial revolution, the accumulation of greenhouse gases in the atmosphere has been rapidly increasing, especially; as a result of human activities such as burning of fossil fuels, land use changes, deforestation and industrial processes. With the understanding that these increases in the concentration of greenhouse gases are beginning to pose serious threats, and that this problem can be resolved with global efforts, global scale co-operation and organizations have been started to form. In this process, The "Framework Convention on Climate Change" signed at the United Nations Conference on Environment and Development held in Rio de Janeiro in June 1992 under the auspices of the United Nations and The "Kyoto Protocol" formed as a result of the "Third Party Conference" held in Japan in December 1997 are special important development (Dam, 2014).

One of the main reasons why the Kyoto Protocol is so important is that it obligates the parties with greenhouse gas reductions on certain dates and at certain rates. In addition, the Kyoto Protocol has introduced three new mechanisms that provide flexibility for greenhouse gas reduction (Gençer, 2008). These mechanisms called the Kyoto Protocol Flexibility Mechanisms, allow countries for greenhouse gas reduction to undertake joint activities outside their own borders. Turkey, in 2004, has became a party to the United Nations Framework Convention on Climate Change, signed the protocol

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

in 2009 and has been the 178th of countries that ratified the Kyoto Protocol. Between 2008 and 2012, Turkey had no obligation to any greenhouse gas emission reduction. On the other hand; according to the protocol, Turkey aims to reduce at least 20% reduction in carbon emissions until 2020.

The Kyoto Protocol is the first step in the climate protocol. However, this protocol is not a result. Because, only developed countries took on the greenhouse gas reduction obligation in the Kyoto Protocol. The world was trying to establish a new agreement on the global scale since Kyoto Protocol. In Paris on December 12, 2015, countries adopted an international agreement to address climate change that requires deeper emissions reduction commitments from all countries—developed and developing. The Paris Climate Agreement, in which 196 countries have reached consensus and 187 countries have proposed a reduction plan (INDC), includes countries that account for more than 96% of total global emissions. The objective of the agreement is to maintain the increase in global temperatures well below 2 degrees Celsius above pre-industrial levels, whilst making efforts to limit the increase to 1.5 degrees (Savaşan, 2017).

The agreement aims to ensure global greenhouse gas emissions peak as soon as possible, and to balance emissions and removals of greenhouse gases in the second half of this century. Furthermore, the agreement addresses adaptation to climate change, financial and other support for developing countries, technology transfer and capacity building, as well as loss and damage (Savaşan, 2017).

This new agreement has set countries' minimum obligations, implemented mechanisms to spur additional action in developing countries, supported the most vulnerable countries in addressing climate change, and established systems to hold countries to their commitments. The Paris Agreement will be strengthened over time using its solid framework.

Turkey, in the context of the struggle against climate change, presented the Intended Nationally Determined Contribution (INDC) plan as a volunteer to the UN Secretariat before the Climate Summit in Paris. Turkey stated that aims to increase greenhouse gas emissions less than 21% from the reference scenario until 2030 in this declaration. Turkey which refers to special conditions in INDC Plan, set targets for renewable energy use and the transition to nuclear energy. Also, Turkey declared that greenhouse gas emissions will be reduced by low-carbon investments that will provide energy efficiency in urbanization, transport and industry sector (Karakaya, 2015).

The atmosphere, which is of vital importance to the living creatures in the world, contains many gases in it. Nitrogen (78.08%) and oxygen (20.95%) in the atmosphere form 99% of the volume of clean and dry air. The remaining dry air section consists of argon (0.93%), an inert gas, and some trace gases, which are very small in quantity. Although its deposition in the atmosphere is very small, the CO_2, which is an important greenhouse gas, is in fourth place with 0.03% (Türkeş, 2008). Carbon dioxide (CO_2) is the most important greenhouse gas that is effective on global warming. Although it is at a very low volumetric rate in the atmosphere, its share in total greenhouse gases exceeds 80% (Koçak, 2006). As a result of from decay of plants, volcanic activity, human and animal breath and also combustion of fossil fuels such as coal, oil and natural gas, CO_2 accumulates in the atmosphere. Because of burning fossil fuels, forest fires, etc., the CO_2 concentration has been increasing rapidly since the beginning of this century. While CO_2 concentration is 290 ppm in the early 20th century, it is estimated that CO_2 concentration will increase to 500 ppm at the end of the 21st century. CO_2 absorbs some of the infrared heat energy released from the earth. Therefore, an increase in CO_2

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

concentration will cause an increase in the average surface temperature of the world scale (Koçak, 2006).

Turkey's carbon dioxide emissions data table and chart are given in *Table 1* and *Figure 1* between 1990 and 2015.

Table 1. *Greenhouse gas emissions by sectors (CO_2 equivalent), 1990-2015. (Source: TUİK (2015): Seragazı emisyon Envanteri)*

Year	Total	Change compared to 1990 (%)	Energy	Industrial processes and product use	Agriculture	Waste
1990	214.0	-	134.4	23.7	44.8	11.1
1991	221.1	3.3	138.5	25.4	45.8	11.3
1992	227.4	6.3	144.7	25.1	46.1	11.5
1993	236.7	10.6	152.2	26.0	46.8	11.8
1994	230.3	7.6	148.9	25.3	44.0	12.0
1995	246.6	15.2	163.5	27.3	43.4	12.4
1996	264.2	23.5	179.2	28.1	44.2	12.7
1997	275.6	28.8	191.2	29.0	42.1	13.2
1998	277.6	29.7	191.0	29.3	43.7	13.5
1999	276.4	29.2	190.2	27.8	44.4	14.0
2000	296.5	38.6	211.7	27.8	42.5	14.5
2001	277.7	29.8	195.0	27.9	39.8	15.0
2002	284.6	33.0	201.9	29.3	38.0	15.4
2003	304.1	42.1	216.6	30.5	41.2	15.9
2004	315.1	47.3	223.3	33.1	42.2	16.5
2005	337.2	57.6	241.0	35.9	43.3	16.9
2006	361.7	69.0	260.5	39.0	44.8	17.5
2007	395.0	84.6	291.4	41.5	44.4	17.7
2008	391.8	83.1	288.5	43.4	42.1	17.8
2009	400.9	87.4	294.6	45.1	43.4	17.9
2010	406.8	90.1	291.8	51.0	45.8	18.2
2011	436.4	103.9	313.9	55.8	48.1	18.5
2012	448.9	109.8	319.3	57.7	53.8	18.1
2013	442.2	106.6	308.3	60.2	57.2	16.5
2014	455.6	112.9	321.2	60.8	57.2	16.4
2015	475.1	122.0	340.0	60.7	57.4	16.9

According to the *Table 1*, the total greenhouse emissions of 2015 as CO_2 equivalents increased 122% compared to 1990. Total greenhouse gas emissions in 2015 were calculated as 475.1 million tonnes (Mt) CO_2 equivalent. The largest share of the CO_2 equivalent in 2015 emissions was energy-related emissions with 71.6% followed by industrial operations and product use by 12.8%, agricultural activities by 12.1% and waste by 3.5%. There are many studies on climate change and greenhouse gas emissions. In this study, we tried to classify in terms of CO_2 emissions the Organisation for Economic Co-operation and Development (OECD) countries which also Turkey is a member. For this purpose, we used fuzzy c-means algorithm which is most commonly

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

known in fuzzy partition. Then, in order to make a comparison, analyzes were made by classical k-medoid algorithm. For analysis of greenhouse gas emissions, which is one of the important ecological issues, it is aimed to be shown that besides classical techniques, fuzzy techniques which are more flexible can be used.

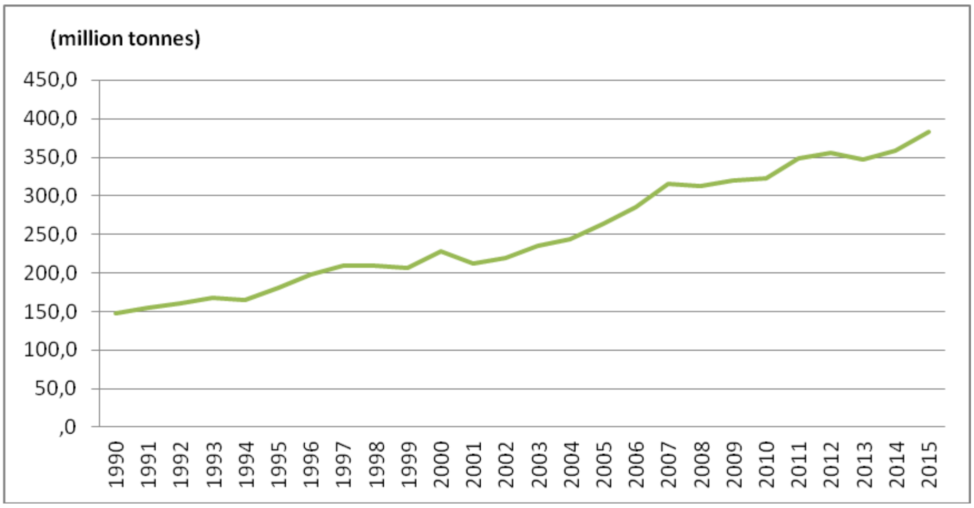

Figure 1. *CO_2 emissions in Turkey between 1990 and 2015*

Methods and materials

Clustering

Clustering analysis or simply clustering is one of the multivariate statistical methods helps to divide data labels of which are not exactly known to subgroups according to similarities and explore different correlation and structures in large data sets. The desired goal in this analysis is to divide the data set so that the elements assigned to the same cluster are as similar as possible but that the two objects in different clusters should not be as similar as possible. Hence, clustering analysis tries to model the ability in clustering similar objects into homogeneous classes or categories. In clustering, initially the number of clusters and number of object that assigned to clusters is not certain. By collecting similar objects into clusters, clustering analysis tries to recreate unknown clusters, hoping that each cluster found represents an actual object category or category. Classifying datasets, clustering analysis only uses some mathematical criteria on the composition of clusters. For this reason, clustering algorithms are equipped with distance measures to measure the similarity of the presented samples.

Clustering techniques are distinguished according to how assignments are made, in other words, according to their assignment patterns (Hartigan, 1975). In classical cluster analysis, each datum must be exactly assigned to a cluster. These classical methods convert non-empty partitions of the data into non-empty and double disjoint subclasses. Such hard data aggregation way can be insufficient when data is equal distance to two or more clusters. Such special data points may represent similar hybrid or mixed objects at a rate equal to two or more species. Classical (hard) partition forces these data points to be fully assigned to one of the clusters; but they should belong to all at equal rates. Hard partition is defined as below;

Let $X = \{x_1, x_2, \dots x_n\}$be a set of n vectors in R^d, representing the data. For an integer $c > 2$, a hard clustering of X into c clusters consists of c disjoint subsets of X', $S_1, \dots, S_c$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

whose union is X. An equivalent way of defining clusters is achieved by using functions. For each $i = 1, \ldots, c$ define $u_i: X \to \{0,1\}$ by $u_i(x) = 1$ if $x \in S_i$ and $u_i(x) = 0$ if $x \notin S_i$ (Bezdek, 1973).

These functions are called membership functions. Because it determines each data point belongs a specific cluster. In hard clustering, membership value is equal to 1 if the data belongs to the cluster, and 0 otherwise. As can be seen, in classical clustering theory, boundaries of clusters and adjectives of clusters are definite. In real life, however, the boundaries of each cluster and the attributes of each element belonging to these clusters are not so certain. Thus, it can be seen that the concept of classical clustering is inadequate in some situations encountered in real life.

Classical clustering analysis is organized under hierarchical and non-hierarchical headings. Hierarchical methods generate arrays of clusters usually starting from clusters that contain a single unit, by performing concatenation until all units are aggregated in a cluster or they generate clusters by starting from cluster that contain all units and separating units sequentially. Non-hierarchical techniques are methods which aim to divide units into clusters that are homogeneous in themselves and heterogeneous among themselves and estimate the parameters of sub-populations (group or cluster mean vectors and covariance matrices) through prototype sets. The most common non-hierarchical methods are k-means clustering, k-medoid clustering and fuzzy clustering. In this study we will study k-medoid and fuzzy methods.

K-medoids algorithms

It attempts to reduce the distance between a point labeled on the cluster and center of cluster. The method selects medoid, which means point the closest to center of cluster, instead of the average of data as a cluster center. Medoids for each cluster are calculated by using the following formula (*Eq. 1*) by finding the object i within the cluster that minimizes

$$\sum_{j \in C_i} d(i, j) \tag{Eq.1}$$

where c_i is the cluster containing object i and $d(i, j)$ is the distance between objects i and j.

There are two advantages of using existing rules as the centers of the clusters. First, a medoid rule serves to usefully describe the cluster. Second, there is no need for repeated calculation of distances at each iteration, since the k-medoids algorithm can simply look up distances from a distance matrix. The most common realization of *k*-medoid clustering is the partitioning around medoids algorithm and is as follows (Kaufman and Rousseeuw, 1990):

- Initialize: randomly select *k* of the *n* data points as the medoids.
- Assignment step: associate each data point to the closest medoid.
- Update step: for each medoid *m* and each data point *o* associated to *m* swap *m* and *o* and compute the total cost of the configuration (that is, the average dissimilarity of *o* to all the data points associated to *m*). Select the medoid *o* with the lowest cost of the configuration. Repeat alternating steps 2 and 3 until there is no change in the assignments.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

Kaufman and Rousseeuw suggested the use of silhouettes both to determine which objects lie well within their clusters and which do not and also to judge the quality of the clustering obtained. For each object i, let $a(i)$ be the average distance of i from all other objects in cluster c_i . For every other cluster $c \neq c_i$, let $d(i,c)$ be the average distance of i from the objects in c. After computing $d(i,c)$ for all clusters $c \neq c_i$, let $b(i)$ be the smallest. The cluster for which this minimum is attained is called the neighbour of i. The number s(i) is given by using *Equation 2:*

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad \text{(Eq.2)}$$

If $s(i)$ is close to 1, object i can be said to be 'well classified'. If $s(i)$ is close to 0, it is unclear whether i should belong to cluster c_i or to its neighbor. A negative value suggests that i has been misclassified. The following summary values are also defined:

- The mean of $s(i)$ for all objects i in a cluster is called the average silhouette width of that cluster.
- The mean of $s(i)$ for all the objects is called the average silhouette width for the entire data set and is denoted by $\bar{s}(c)$, where c is the number of clusters.

Kaufman and Rousseeuw suggest that $\bar{s}(c)$, can be used for the selection of a best value of c, by choosing that k for which $\bar{s}(c)$, is maximal (Richards and Rayward-Smith, 2001).

Fuzzy clustering

Fuzzy clustering is the generalization of this point of view. In fuzzy set theory, an element belongs to a cluster with membership values between 0 and 1, including 0 and 1. In other words, in a fuzzy set, a fuzzy set element belongs partially to a cluster.

A fuzzy clustering of X into c clusters consist of functions $u_1, \dots, u_c$ where $u_i: X \to [0,1]$ and $X_i u_i(x) = 1$, for all $x \in X$. Formally, given a set of objects, $\{x_1, x_2, \dots x_n\}$ a fuzzy clustering of *k* fuzzy clusters, $c_1, \dots, c_k$, can be represented using a partition matrix, $U = [u_{ij}].(1 \leq i \leq n, 1 \leq j \leq k)$, where u_{ij} is the membership degree of x_i in fuzzy cluster c_j. The partition matrix should satisfy the following three requirements:

- For each object, x_i, and cluster, c_j, $0 \leq u_{ij} \leq 1$. This requirement enforces that a fuzzy cluster is a fuzzy set.
- For each object, x_i, $\sum_{j=1}^{k} u_{ij} = 1$. This requirement ensures that every object participates in the clustering equivalently.
- For each cluster, c_j, $0 < \sum_{i=1}^{n} u_{ij} < n$. This requirement ensures that for every cluster, there is at least one object for which the membership value is nonzero.

Fuzzy cluster analysis therefore allows gradual memberships of data points to clusters in [0, 1]. This gives the flexibility to express that data points belong to more than one cluster at the same time. Furthermore, these membership degrees offer a much finer degree of detail of the data model. Aside from assigning a data point to clusters in (equal) shares, membership degrees can also express how ambiguously or definitely a data point should belong to a cluster.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

The first use of this method based on fuzzy set theory in clustering analysis was carried out by Bellman et al. (1966). Many fuzzy clustering algorithms are based on the optimization of the objective function. The most widely used algorithm is the fuzzy c-means algorithm (Ruspini, 1970).

Fuzzy c-means algorithms

The fuzzy c-means algorithm is presented in its initial form by Dunn (1974) as an alternative to the classical k-means clusters and is completed by Bezdek (1974). The fuzzy c-means algorithm is based on the objective function. A large family of fuzzy clustering algorithms have been created to attempt to minimize the objective function of these basic c-means algorithms. In short, the most basic algorithm for fuzzy clustering analysis is the fuzzy c-means algorithm. The mentioned objective function is defined as *Equation 3* (Bezdek, 1981):

$$J(X;U,V) = \sum_{j=1}^{c}\sum_{i=1}^{n}\left(u_{ji}\right)^{m}\left\|x_i - v_j\right\|^2 \quad \text{(Eq.3)}$$

where n is the total number of patterns in a given data set and c is the number of clusters;$X = \{x_1, \dots x_n\} \subset R$ and $V = \{v_1, \dots, v_c\} \subset R$ are the feature data and cluster centroids; $U = [u_{ji}]$ and $c \times n$ is a fuzzy partition matrix composed of the membership grade of pattern xi to each cluster j. value is the total of pattern measurement of all weighted least square errors. The weighting exponent m is called the being effective on the clustering performance of FCM (Şahinli, 1999).

Unfortunately, the object function $J\,(X;U,V)$ can not be directly minimized. Therefore, an iterative algorithm is used, which alternately optimizes the membership degrees and the cluster parameters. That is, first, the membership degrees are optimized by using *Equaition 4* for fixed cluster parameters, then the cluster parameters are optimized by using *Equation 5* for fixed membership degrees:

$$U_t = J_U\left(V_{t-1}\right) \quad \text{(Eq.4)}$$

and

$$V_t = J_V\left(U_t\right) \quad \text{(Eq.5)}$$

The main advantage of this scheme is that the optimum can be computed directly in each of two steps. By iterating the two steps, the joint optimum is approached (although it cannot be guaranteed that the global optimum will be reached—the algorithm may get stuck in a local minimum of the objective function *J*).

The update formulae J_U and J_V are derived by simply setting the derivative of the objective function J w.r.t. the parameters to optimize equal to zero (Höppner et al., 1999). Independent of the chosen distance measure the following update formula (*Eq. 6*) for the membership degree and following update formula (*Eq. 7*) for prototype is obtained from J(X;U,V):

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

$$u_{ji} = \left[\frac{1}{\sum_{k=1}^{c} \left[\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right]^{2/m-1}} \right], 1 \leq j \leq c, 1 \leq i \leq n \quad \text{(Eq.6)}$$

$$v_j = \frac{\sum_{i=1}^{n} (u_{ji})^m x_i}{\sum_{i=1}^{n} (u_{ji})^m}, \quad 1 \leq j \leq c \quad \text{(Eq.7)}$$

In *Table 2* a brief description of the fuzzy *c*-means algorithm is given. Algorithm can be terminated when the relative change in the cluster centers gets very small or the objective function *J* cannot be minimized anymore (Bezdek, 1974).

Table 2. *FCM algorithm steps*

- Select the fuzzifier exponent m ($m > 1$) and initialize the fuzzy partition matrix $U = (u_{ji})$ randomly
- **while** termination conditions not met **do**
1. Compute the cluster centers v_j
2. Update the fuzzy partition matrix $U = (u_{ji})$
end while

The fuzzy c-means algorithm detects clusters in the form of dots. It is not effective in finding other cluster shapes. Also, the fuzzy c-means algorithm has some advantages and disadvantages. Sensitivity to noise points, adherence to initial values, long-running computation time are the disadvantages of the algorithm. On the other hand, unsupervised and constant convergence are the main advantages of the algorithm.

As seen in the first step of the fuzzy c-means algorithms, determining of the number of initial clusters is an important task. The fact that the number of the initial cluster is not known in advance makes it necessary to test the validity of the clusters. The validity indices are used to determine which set of clusters is appropriate clustering from the clusters in the fuzzy clusters. There are many validity indices in the literature. The most commonly used index in fuzzy clustering analysis is the Xie-Beni index. The Xie-Beni index is defined as follows in *Equation 8* (Xie and Beni, 1991):

$$XB(c) = \frac{\sum_{j=1}^{c} \sum_{i=1}^{n} (u_{ji})^m \|x_i - v_j\|^2}{N \min_{ji} \|x_i - v_j\|^2} \quad \text{(Eq.8)}$$

Xie-Beni index is also known as the density and separation index. It is based on two basic features; compactness and separation. In *Equation 8,* the numerator shows the closeness of the fuzzy partition, i.e. the closeness of the elements in the cluster, and the denominator shows the strong separation between the clusters, i.e. the distance between

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

the clusters (Xie and Beni, 1991). It is seen in empirical studies that this index should take small values.

Results and discussion

In this study, it is aimed to determine the classification of the OECD countries by fuzzy clustering and classical clustering analysis in terms of the indicators from fossil fuel consumption and detect the cluster Turkey belongs, and other countries in this cluster, and evaluate whether it shows similar functions between Turkey and these countries or not. For the analysis, we used CO_2 emissions data from fuel combustion by sector from open access platform -International Energy Agency (IEA). 32 countries and five variables are used. The variables used are shown in *Table 3*.

Table 3. *Used variables*

Variables ID	Variables
X_1	Electricity and heat production
X_2	Other energy industry own use
X_3	Manuf. industries and construction
X_4	Transport
X_5	Other sectors

For the application, R studio 3.4.0 which is an environment for statistical computing program and “fclust” package was used. In order to apply the algorithm, the initial parameters (excluding the number of cluster) were selected randomly. The Xie-Beni results were calculated to determine the number of clusters.

As seen in *Table 4* and *Figure 2*, the experimental results revealed that for fuzzy c - mean algorithms, the elbow was located at c = 4. On the other hand, the smallest value of Xie-Beni index is the number of cluster is 4.

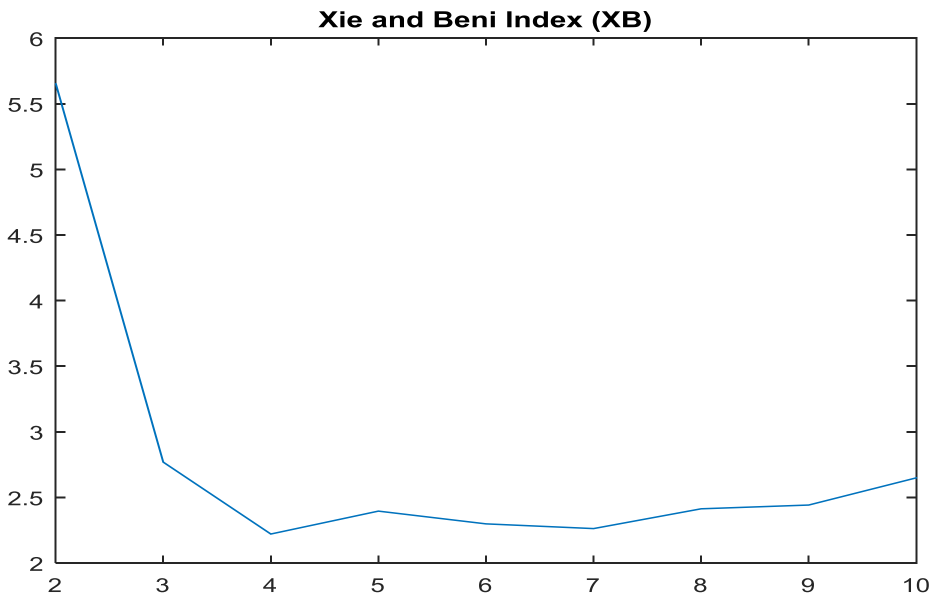

Figure 2. *Emissions cluster for OECD Countries, Xie- Beni graphic for number of cluster*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

Table 4. *Results of Xie-Beni index*

c	2	3	4	5	6	7	8	9	10
XB	***5.657***	***2.769***	***2.220***	***2.395***	***2.298***	***2.262***	***2.413***	***2.441***	***2.649***

Scattering of the clusters which are obtained from result of fuzzy c- means algorithm in Rstudio package program are shown as in *Figure 3*.

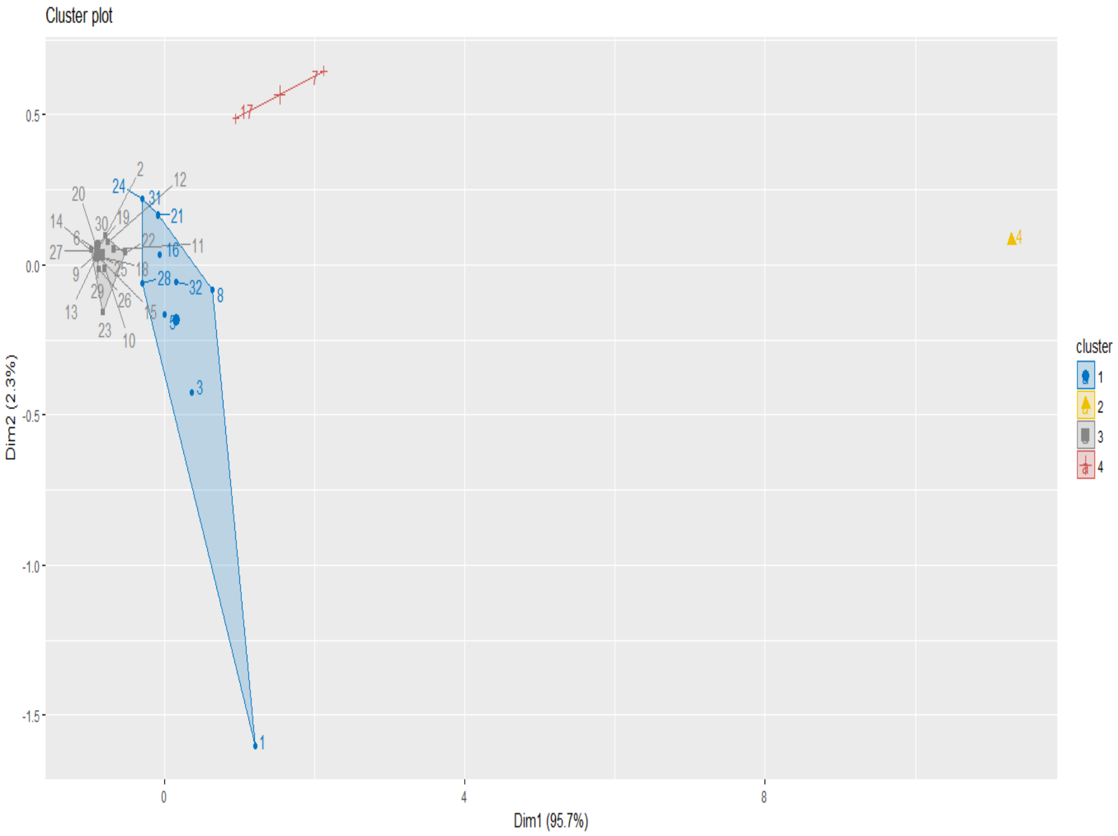

Figure 3. *Emissions cluster for OECD countries, c = 4*

Membership values of OECD countries are given in *Table 5*. If a unit has the largest membership value in a cluster, then it is assigned to this cluster.

Table 5. *Cluster membership degrees for OECD member countries*

Countries	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
CND (Canada)	0.001	0.255	0.611	0.133
CL (Chile)	0.999	0.000	0.000	0.001
MEX (Mexico)	0.000	0.005	0.974	0.021
USA (United States)	0.000	1.000	0.000	0.000
AUS (Australia)	0.000	0.002	0.935	0.063
ISR (Israel)	0.999	0.000	0.000	0.001

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

J (Japan)	0.000	0.000	0.001	0.999
KOR (South Korea**)**	0.000	0.019	0.962	0.019
NZ (New Zealand)	0.999	0.000	0.000	0.001
A (Austria)	0.999	0.000	0.000	0.001
B (Belgium)	0.998	0.000	0.000	0.002
CS (Czech Republic)	0.999	0.000	0.000	0.001
DK (Denmark**)**	0.999	0.000	0.001	0.000
EST (Estonia)	0.996	0.001	0.000	0.003
FIN (Finland)	0.999	0.001	0.000	0.000
F (France)	0.000	0.002	0.849	0.149
G (Germany)	0.001	0.156	0.054	0.789
GR (Greece)	0.999	0.000	0.000	0.001
H (Hungary)	0.999	0.000	0.000	0.001
IRL (Ireland)	0.990	0.000	0.001	0.009
I (Italy)	0.000	0.002	0.788	0.209
NL (Netherlands)	0.982	0.000	0.000	0.018
N (Norway)	0.998	0.000	0.000	0.002
PL (Poland)	0.000	0.002	0.694	0.304
P (Portugal)	0.990	0.003	0.001	0.006
SK (Slovakia)	0.983	0.002	0.000	0.015
SVN (Slovenia)	0.996	0.002	0.000	0.002
E (Spain)	0.000	0.001	0.651	0.348
S (Sweden)	0.999	0.000	0.000	0.001
CH (Switzerland)	0.999	0.000	0.000	0.001
TR (Turkey)	0.000	0.002	0.872	0.126
GB (Great Britain)	0.000	0.001	0.986	0.013

As a result of the analysis, with the help of fuzzy c-means clustering, the number of cluster has been determined as 4. As can be seen in *Table 6*, the first cluster was determined as New Zealand, Chile, Austria, Belgium, Czech Republic, Denmark, Israel, Finland, Greece, Hungary, Ireland, Netherlands, Norway, Poland, Portugal, Slovakia, Sweden, Switzerland. While United States took place in the second cluster, the third cluster were identified as Canada, Mexico, Australia, South Korea, France, Italy, Poland, Spain, Turkey and Great Britain. The countries in the last group were designated as Japan and Germany.

In order to compare the results of fuzzy clustering, the k-medoid clustering method of classical clustering methods is also used. NCSS program was used for analysis. After data entry in NCSS for k-medoid application, “Medoid Partitioning” was selected from the “Clustering” section in the analysis section and the number of clusters was entered as minimum 2 and maximum 10 in order to determine the optimal number of clusters. The results are shown in *Table 7*.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

Table 6. *Clusters according to the countries(FCM)*

Countries	Clusters
ISR CL NZ A B CS DK EST FIN GR H IRL NL N P SK SVN S CH	1
USA	2
CND MEX AUS KOR F I PL E TR GB	3
J G	4

Table 7. *Silhouette statistics according to the number of cluster*

Number Clusters	Minimize This	Adjusted	Maximize This
	Average Distance	Average Distance	Average Silhouette
2	**189.034**	**11.815**	**0.890**
3	48.701	4.566	0.530
4	24.173	3.022	0.543
5	27.745	4.335	0.568
6	18.111	3.396	0.563
7	15.991	3.498	0.570
8	12.068	3.017	0.507
9	9.822	2.762	0.439
10	9.581	2.994	0.429

Based on the average silhouette statistics obtained for the clusters, the number of clusters with which the average silhouette statistics value is maximized as k = 2.

According to k-medoid clustering, OECD countries were divided into 2 clusters in points of carbon dioxide emissions. As seen in *Table 8,* the classification was made in the United States as a cluster, the remaining 31 OECD countries as the second cluster.

Table 8. *Clusters according to the countries(K-medoid)*

Cluster 1	**USA**
Cluster 2	**CND CL MEX AUS ISR J KOR NZ A B CS DK EST FIN F G GR H NL N PL P SK SVN E S CH TR GB**

Conclusions

Societies are increasingly fragile against atmosphere-related natural disasters. As a result of increasing greenhouse gas emissions due to human activities, the link between industry-climate change is getting stronger. Although the kyoto protocol is the basic protocol, the results have not been effective. For this reason, the Paris Climate Agreement, which is one of the important protocols on fighting climate change, was organized in 2015. The main objective of the protocol is to reduce the emissions of all greenhouse gases, primarily carbon dioxide. For each country, the measures are clearly stated in the protocol.

The changes in the atmosphere of the earth largely took place in the years following the industrial revolution and especially in the 20th century. In the 21st century, it is estimated that greenhouse gas emissions will further increase the global average temperature. Carbon dioxide (CO_2), the most important of the six greenhouse gases, its emission have increased especially during the past century. In this study, CO_2 emissions resulting from the fuel combustion of 32 countries were classified by using fuzzy c-means algorithm - based on fuzzy logic-, and classical k-medoid algorithm. As a result of the analysis, in the fuzzy clustering analysis, the cluster number is 4; whereas in classical clustering analysis, the number of clusters was found to 2.

It can be said that the k-medoid clustering analysis is created an artificial cluster because the first cluster is only formed by the USA. It can also be said that the classification based on the fuzzy c-means algorithm reflects the natural classification resulting from the combination of the variables in the data matrix. Since the data used in the study show a global structure, the fuzzy c-means method of conventional fuzzy clustering methods was used and produced useful results.

Japan and Germany, being in fourth cluster in the clustering result, are ranked 2nd and 3rd, respectively, according to CO_2 emission values for 2015. According to the year 2015 statistics, the total CO_2 emission was 32.294 million metric tons in the world, while the emission value of the US was 4997 million metric tons. As for the analyzed CO_2 emission variables of the USA, clustering of the US in a single cluster is an important and verifiable result in this matter.

According to the 2015 greenhouse emission inventory report, published by the Turkish Statistical Institute, in Turkey, total greenhouse gas emissions in 2015 were 475.1 million tons (mt), equivalent to CO_2. In this inventory, the largest share of CO_2 emissions for 2015 created energy-driven emissions.

CO_2 emissions in the world has being estimated to increase by an average of 2% each year since 1971 and by 45% until 2030. With this study, it is aimed to shed light on researchers working on environment and climate change. However, classification for carbon dioxide emission can be enriched by adding other countries besides OECD countries or different greenhouse gas statistics can be clustered for countries. Thus, greenhouse gas emissions of countries can be scrutinized from a wider perspective.

REFERENCES

[1] Bellman, R., Kalaba, R., Zadeh, L. A. (1966): Abstraction and pattern classification. – Journal of Mathematical Analysis and Applications 13: 1-7.

[2] Bezdek, J. C. (1973): Fuzzy Mathematics in Pattern Classification. – Ph.D. Dissertation, Cornell University, Ithaca, NY.

[3] Bezdek, J. C. (1974): Cluster Validity with Fuzzy Sets. – J. Cybernet 3: 58-73.

[4] Bezdek, J. C. (1981): Pattern Recognition with Fuzzy Objective Function Algorithms. – Plenum Press, NewYork.

[5] Dam, M. M. (2014): Sera Gazi Emisyonlarinin Makroekonomik Değişkenlerle İlişkisi: Oecd Ülkeleri için Panel Veri Analizi. – Doktora Tezi, İktisat, Adnan Menderes Üniversitesi, Aydın.

[6] Dunn, J. C. (1974): A fuzzy relative of ISODATA process and its use in detecting compact, well separated clusters. – Journ. Cybern. 3: 95-104.

[7] Gençer, B. B. (2008): Değişen İklimler, Enerji ve Türkiye. – Araştırma Notu, Ekonomik ve Toplumsal Araştırmalar Merkezi (BETAM), Bahçeşehir Üniversitesi, İstanbul.

[8] Hartigan, J. A. (1975): Clustering Algorithms. – Wiley, NewYork.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526

[9] Höppner, F., Klawonn, F., Rudolf, K., Runkler, T. (1999): Fuzzy Cluster Analysis: Methods for Classification Data Analysis and Image Recognition, pp. 5-75. – Wiley, Chichester.
[10] Karakaya, E. (2015): İklim Değişikliği Müzakerelerine Bir Bakış: 2015 Paris İklim Zirvesi. – Journal of Institute of Social Sciences 3: 1-12.
[11] Kaufman, L., Rousseuw, P. (1990): Finding Groups Data: An Introduction to Cluster Analysis. – John Wiley and Sons, New York.
[12] Koçak, K. (2006): İklim Değişiminde İnsan Faktörü. – İstanbul Teknik Üniversitesi. http://web.itu.edu.tr/~kkocak/iklimpdf.pdf.
[13] Richards, G., Rayward-Smith, V. J. (2001): Discovery of association rules in tabular data. – In: Proceedings of IEEE First International Conference on Data Mining, Nov. 29 2001-Dec. 2 2001, pp. 465-473.
[14] Ruspini, E. H. (1970): Numerical methods for fuzzy clustering. – Information Sciences 2: 319-350.
[15] Şahinli, F. (1999): Kümeleme Analizine Fuzzy Set Teorisi Yaklaşımı. – Yüksek Lisans Tezi, İstatistik, Gazi Üniversitesi Fen bilimleri Enstitüsü, Ankara.
[16] Savaşan, Z. (2017): A brief assessment on the Paris Climate Agreement and compliance issue. – Uluslararası İliskiler 14(54): 107-125.
[17] TUİK (2015): Seragazı emisyon Envanteri, www.tuik.gov.tr, internet search.
[18] Türkeş, M. (2003): Sera Gazı Salımlarının Azaltılması İçin Sürdürülebilir Teknolojik ve Davranışsal Seçenekler. – V.Ulusal Çevre Mühendisliği Kongresi: Çevre Bilim ve Teknoloji Küreselleşmenin Yansımaları, Bildiriler Kitabı, pp. 267-285.
[19] Türkeş, M. (2008): İklim Değişikliğiyle Savaşım, Kyoto Protokolü ve Türkiye. – Mülkiye 259: 101-131.
[20] Xie, L., Beni, G. (1991): A validity measure for fuzzy clustering. – IEEE Trans PAMI 13: 841-847.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1603_25132526