HIERARCHICAL CLUSTERING BASED ON INTERNATIONAL SUSTAINABILITY INDICES OF EU COUNTRIES

RADKA REPISKÁ[1] ─ NORA GRISÁKOVÁ[2] ─ PETER ŠTETKA[3]

Abstract: *The presented paper focuses on the possibility to group countries by the cluster method in terms of assessing the sustainable competitiveness of European countries. Our calculation is based on HDI (Human Development Index) and EPI (Environmental Performance Index) indices. We also tried to show the differences in HDI and EPI index of the Slovak Republic and the Netherlands. The aim of this paper is to evaluate the global competitiveness regarding the environmental economics model, considering all three levels: economic, social, and environmental. We measure the socio-economic dimension using HDI according to the health and education areas, then we measure the environmental dimension using EPI, which monitors the behaviour of countries in the field of human health protection and ecosystem protection. Our question is whether there is an appropriate classification for the development of these countries that could help to reduce the differences between the average countries and the EU 27 average. The approach to this topic began with the question whether these countries, which have high values of economic growth, have a high level of EPI or HDI. The intention is to look for the possible existence of a gradual rapprochement of countries belonging to the same group.*

Keywords: *Cluster Analysis, Sustainability indices, HDI Index, EPI Index, Slovak Republic, Netherlands, SAS Software*

JEL Classification: C38, I00, P28, Q56

[1] Ing. Radka Repiská, University of Economics in Bratislava, Slovak Republic,
e-mail: radka.repiska@euba.sk, https://orcid.org/0000-0002-2192-9316
[2] doc. Ing. Nora Grisáková, PhD., University of Economics in Bratislava, Slovak Republic,
e-mail: nora.grisakova@euba.sk, https://orcid.org/0000-0001-7032-042X
[3] Ing. Peter Štetka, PhD. University of Economics in Bratislava, Slovak Republic,
e-mail: peter.stetka@euba.sk, https://orcid.org/0000-0001-8444-3332

EKONOMICKÉ ROZHĽADY – ECONOMIC REVIEW, 2022, 51(2), 149 – 170
https://doi.org/10.53465/ER.2644-7185.2022.2.149-170

1 Introduction

In today's world, which is set for constant economic growth and a highly consumerist way of life in society, when demand is creating another demand, we are increasingly addressing the question of what to do and how to proceed in order to leave the society in the best possible condition for future generations and how to endanger the environment in the least way. In recent times we often meet with concepts such as climate change, global warming, European Green Deal ("Green Deal"), sustainable development and variety of modern society's topics and environmental economics.

The aim of this paper is to evaluate the global competitiveness regarding the environmental economics model, considering all three levels: economic, social, and environmental. We measure the socio-economic dimension using HDI according to the health and education areas, then we measure the environmental dimension using EPI, which monitors the behavior of countries in the field of human health protection and ecosystem protection. This paper focuses on the possibility to group countries by the cluster method in terms of assessing the sustainable competitiveness of European countries, especially Slovakia and the Netherlands. The question is whether there is an appropriate classification for the development of these countries that could help to reduce the differences between the average countries and the EU 27 average. The approach to this topic began with the question whether these countries, which have high values of economic growth, have a high level of EPI or HDI. The intention is to look for the possible existence of a gradual rapprochement of countries belonging to the same group.

We decided to examine 2 European countries - the Slovak Republic and the Netherlands. A country is not a business, therefore, it is more difficult to focus on the most important things which could emerge greatly from our research. The reason why we decided to compare these countries is their very similar distribution in terms of area. Slovakia is our home country and the Netherlands as a European exchange country helped us to answer several questions, particularly in key aspects of sustainable development of our plans as well as the issue that even a small country can be successful and environment friendly and prevent or combat the threats of climate change.

In 1939, Tryon (1939) first used the term of cluster analysis. Cluster analysis is a classification procedure that groups objects into distinct subgroups that are similar within but different than objects included in other subgroups. The resulting branching diagram is a classification that provides a sequence of clusters (subgroups) according to which a group of objects is divided. For instance, if several ecological units are examined, this analysis is suitable for showing species composition patterns between these units. Cluster analysis essentially creates a dendrogram or tree, the branches of which represent each of the ecological units, and the data on the species composition of these places determine the structure of the branch. Merged branches represent groups or clusters of sites with a similar species composition and the length of a branch before merging is inversely proportional to the degree of similarity of the species composition.

There is a wide range of cluster analyses, we focused on hierarchical, agglomerative, where each object is considered a cluster. The choice of an appropriate method is crucial because it determines (partially) a classification derived from species composition data. Like many multidimensional statistical analyses, cluster analysis attempts to represent complex relationships between objects, in our case between countries, in a simple one-dimensional way. We processed the application of cluster analysis using a comparison of 3 classifications on a set of 15 EU countries. The status of all acquired variables reflects the observed period of the most recently obtained data at the end of 2018, which represents the full coverage of the variables HDI (Human Development Index) and EPI (Environmental Performance Index) for all monitored countries.

2 Methods and methodology

Cluster analysis of a multidimensional data set aims to divide a large set of data into meaningful subgroups of subjects. In cluster analysis, many methods are available to classify objects based on their (un) similarity (Johnson, 1967). Dasgupta (2016) framed similarity-based hierarchical clustering as a combinatorial optimization problem, where a "good" hierarchical clustering is one that minimizes a particular cost function. Murtagh and Contreas (2011) made a survey of agglomerative hierarchical clustering algorithms and discussed efficient implementations that are available in R and other software environments. They look at hierarchical self-organizing maps, and mixture classifications reviewed grid-based clustering, focusing on hierarchical density-based approaches. Jafarzadegan, Safi-Esfahani and Beheshti (2019)

EKONOMICKÉ ROZHĽADY – ECONOMIC REVIEW, 2022, 51(2), 149 ─ 170
https://doi.org/10.53465/ER.2644-7185.2022.2.149-170

propose a novel method of combining hierarchical clustering approaches based on principle component analysis (PCA). PCA as an aggregator allows considering all elements of the descriptor matrices. In their approach, basic clusters were made and transformed to descriptor matrices. Then, a final matrix was extracted from the descriptor matrices using PCA and dendrogram was constructed from the matrix that was used to summarize the results of the diverse clustering.

We expand the data matrix X of *pxk* type with p objects and k indicators into the set C by means of clustering procedures with all clusters *m*, where the objects of the primary matrix X were grouped. The total number of clusters *m* has the possibility to range from 1 to *p*, while the best situation occurs when we reach the number of clusters smaller than the number of objects (in our case the studied countries) (Van Vark & Howells, 1984).

From the best-known metrics of distances between objects, we chose the *Euclidean distance of objects* for our analysis, which is set by the following equation (Van Vark & Howells, 1984):

$$d\left(X_i, X_J\right) = \sqrt{\sum_{s=1}^{k}(x_{is} - x_{js})^2} \quad (1)$$

Where:

x_{is} is the value of the s-th variable for the i-th object.

x_{is} is the value of the s-th variable for the j-th object.

This distance measurement, which generalises the concept of physical distance in two- or three-dimensional space to multidimensional space, is often referred to as the "Pythagorean distance" and forms the basis for Ward's method.

The main types of analysis are hierarchical clustering procedures divided into:

- **agglomerative** - the decomposition process begins with each cluster that contains exactly one object and continues the decomposition by a suitably selected method until all of them are merged into one cluster;

- **divisive** - the opposite procedure begins with one cluster containing all objects and gradually splits into smaller clusters (Legendre & Legendre, 2012).

Next, we will deal with hierarchical clustering procedures, where there are several different methods used to determine which clusters should be combined at each stage, *Nearest-neighbour clustering method, Median method and Ward's method* were chosen to collect minimised heterogeneity clusters.

The median method is described by the following two equations (Legendre & Legendre, 2012)

1. Nearest-neighbour clustering method (*"Nearest"*)

$$D_1(C_h, C_r) = min\,\{d(X_i, X_j)\} \quad X_i \in C_h, X_j \in C_r \tag{2}$$

Where:

The *"Nearest"* method uses the distance of the nearest elements of the clusters *Ch* and *Cr.*

2. Median method (*"Median"*)[4]

$$D_2(C_h, C_r) = d\,(\overline{X_h}, \overline{X_r}) \quad where \quad \overline{X_h} = \frac{1}{n_h}\sum_{Xi \in C_h} Xi \;\;,\;\; \overline{X_r} = \frac{1}{n_r}\sum_{Xj \in C_r} Xj \tag{3}$$

Where:

The Median clustering method uses the distance between the medians of two clusters and serves as an improvement to *the Centroid Method.*

Ward's method is a correct hierarchical procedure and makes it possible to determine how many groupings should be considered, and its great advantage is the tendency to remove small clusters and form clusters of roughly the same size. The similarity between 2 clusters is the sum of the squares in the clusters summarised in all variables, the proximity between the 2 clusters being defined as the increase in the square root error resulting from the merging of 2 clusters (Han, Kamber & Pei, 2012). In the case of the Ward's method in terms of distance, equation 4 can be formulated in the form of the product of the Euclidean distance of objects between the centre of clusters conditioned to join and the coefficient, based on the size of the cluster (Řezanková, Húsek & Snášel, 2009):

$$D(C_h, C_r) = \frac{n_h n_r}{n_h + n_r} \times d^2(\overline{X_h}, \overline{X_r}) \tag{4}$$

[4] The Median clustering method uses the distance between the centroids of the clusters and serves as an improvement to *the Centroid method.*

Where:

In hierarchical grouping, the sum of squares starts from zero (each point is in its own grouping) and then increases as we merge the clusters. Ward's method keeps this growth as small as possible. Considering two pairs of clusters whose centres are equidistant from each other, the method prefers to merge the smaller ones.

The results of hierarchical clustering can be displayed graphically, using a tree diagram - "dendrogram", which shows all the steps in a hierarchical process, including distances, where clusters combine.

3 International Sustainability Indices

In this part of the paper, we come to specific variables, sustainability indices and their subsequent analysis applied mainly to the Slovak Republic in comparison with Netherlands. In the case of the HDI index, we used 3 main dimensions and related indicators, compared the achieved results and evaluated the occupation of Slovakia and the Netherlands within the evaluated EU countries. We proceeded in a similar way in the case of the EPI index, where we evaluated Slovakia and Netherlands in 24 performance indicators in ten categories of problems related to Environmental Health and Ecosystem Vitality.

3.1 Research objects - Slovakia and the Netherlands

The country is not a business and therefore it is more difficult to focus on the most important things that could best come out of our research. Based on personal experience on a foreign exchange stay in the Netherlands, we were given the opportunity to see what significant differences in environmental protection are visible in comparison with our country Slovakia.

The reason why we decided to compare these countries is their very similar distribution in terms of size. Slovakia is our home country and the Netherlands, as a European exchange country, has helped us to answer several questions, especially in key aspects of sustainable development of our planet and especially in the issue that even a small country can be successful and friendly to protect the environment and prevent or combat the threats of climate change.

3.2 HDI index and its components in Slovakia and the Netherlands

For the comprehensive part and interconnection of the analysis of environmental indicators, we decided to use one of the most comprehensive indicators of sustainable development of socio-economic nature - HDI ("Human Development Index"). Since 1990, HDI has three dimensions: a long and healthy life, education and a decent standard of living. Four indicators are currently used to capture the three dimensions: life expectancy at birth (long and healthy life); average years of adult education aged 25 and over, expected years of schooling (education); and gross domestic product (GDP) per capita adjusted by purchasing power parity (standard of living).

The HDI calculation process takes place in the following steps:

1. The first step in calculating the HDI is to create three separate indices for each of the these three dimensions, which are then used to calculate the global HDI. In each of the three three-dimensional indices, the country's results are normalized to a score between 0 and 1 using minimum and maximum values.

2. In the second step, the three indices are aggregated to form a global HDI. For this purpose, the 3 dimensional indices are multiplied together and the third root is taken. This gives the geometric mean of the dimension indices.

We rank the HDI index among the aggregated indicators that measure the progress of society in 3 dimensions, which relate to the health, education and living standards of the population. The main principle of the HDI index is to measure the achieved level of human development using a single index and to ensure a comparison of all countries in the world. We calculate the HDI by the geometric mean of 3 determined indices, which we calculated from 3 main dimensions.

The international ranking of countries according to the HDI index is recalculated each year based on the most up-to-date comparable data in the area of 3 dimensions of the human development index: long and healthy life, education and standard of living. The countries are divided according to the size of the HDI index into 4 groups of countries with 4 levels: very high level of HDI, high level of HDI, medium level of HDI and low level of human development.

https://doi.org/10.53465/ER.2644-7185.2022.2.149-170

Table 1: HDI and its components in Slovakia and the Netherlands in 2018

Indicator	Netherlands	Slovakia
HUMAN DEVELOPMENT INDEX	0.934	0.855
LONG AND HEALTHY LIFE		
Life expectancy	82.3	77
EDUCATION		
Estimated number of years of education	12.2	15.0
Average years of schooling	18.0	12.5
STANDARD OF LIVING		
Gross national income per capita (PPP)	46.711	29.467

Source: Own processing according to the United Nation Development Programme (Human development reports, 2020) database.

In Table 1 we compared the Human Development Index of Slovakia and the Netherlands, according to which we can conclude that the Netherlands is one of the countries with a very high level of human development of 0.934 and Slovakia is one of the countries with a high level of HDI of 0.855. Using Table 2 we pointed out how the Human Development Index developed in Slovakia and the Netherlands in the observed period 2000 - 2018.

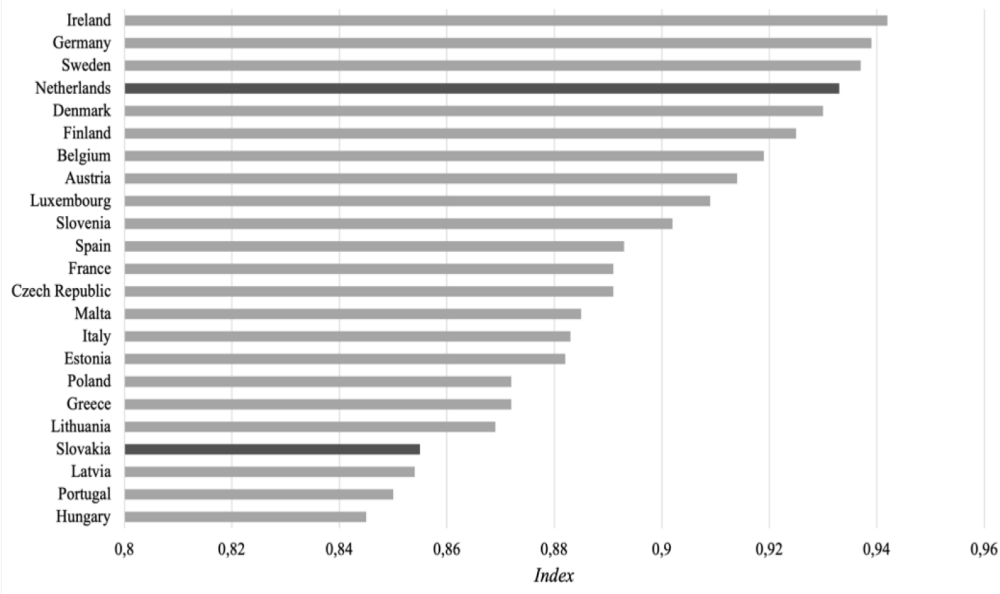
Table 2: Trends in the development of HDI in Slovakia and the Netherlands in 2000 ─ 2018

Years	2000	2010	2011	2012	2013	2014	2015	2016	2017	2018
Netherlands	0.88	0.91	0.921	0.923	0.923	0.924	0.932	0.933	0.934	0.934
Slovakia	0.763	0.827	0.832	0.836	0.839	0.844	0.845	0.853	0.855	0.855

Source: Own processing according to the United Nation Development Programme (Human development reports, 2020) database.

To perform an international comparison of EU countries, we prepared Figure 1 from the data of the Human Development Index of selected 24 member countries drawn from the United Nations Development Program (UNDP). Slovakia took third place among the V4 countries (the Czech Republic, Poland, Slovakia and Hungary) and the Netherlands took 1st place among the economic union of the 3 Benelux countries (The Netherlands, Belgium and Luxembourg).

Figure 1: International Comparison of the Human Development Index (HDI)



Source: Own processing according to UNDP (2020) (United Nations Development Programme: Human Development Index (HDI). Dimension: Composite indices.

3.3 EPI index and its categories in Slovakia and the Netherlands

We chose 2 main indicators for the correct cluster analysis, while we went through the first HDI indicator and in the following we characterize one of the most comprehensive indices - EPI (Environmental Performance Index).

The 2018 Environmental Performance Index (EPI) evaluates 180 countries in 24 performance indicators in ten categories of environmental health problems (40%) and ecosystem vitality (60%). These metrics provide a view on the national scale of how close countries are to achieving environmental policy goals. EPI Framework shows a hierarchy of ten problem categories: Air Quality (26%), Water and Hygiene (12%), Heavy Metals (2%), Biodiversity and Habitat (15%), Forests (6%), Fishing (6%), Climate and Climate (18%), Air Pollution (6%), Water Resources (6%) and Agriculture (3%). These problem categories are then combined into 2 policy objectives - Environmental Health (40%) and Ecosystem Vitality (60%) and finally consolidated into an overall EPI (100%).

The collection process takes place in the following steps:

1. Scores are calculated for each of the ten major policies in the category based on one to four basic indicators, each basic indicator representing a discrete set of data.

2. The scores are further calculated for the environmental objectives and the vitality of the weighted ecosystem.

3. The overall environmental performance index shall then be calculated on the basis of the average of two broad objective scores. The ranking is based on the index score.

A meaningful country-by-country comparison requires a scoreboard for each of the 24 indicators and placed on a common scale, with 0 indicating the worst performance and 100 indicating the best performance. The country's location on this scale determines the extent to which a country achieves its international sustainability goals.

In the following Table 3 we processed the evaluation of the EPI index for the Netherlands and Slovakia. The Netherlands ranked 18th out of 180 countries rated with a score of 75.46 and the best score (EPI 99.90) in the Water Resources category in 3rd place and the worst score (EPI 34.60) in the Fishing category in 131st place Slovakia placed 28th with a score of 70.6, with the best overall EPI from the V4 countries and Eastern European countries. The best score (EPI 75.08) in the category Vitality of ecosystems in 3rd place and the worst score (EPI 59.42) was achieved by Slovakia in the category of air quality, where it was placed in the 133rd position.

Table 3: Evaluation of the EPI index of Slovakia and the Netherlands

Categories	Global Ranking of the Netherlands	EPI Index	Global Ranking of Slovakia	EPI Index
EVALUATION OF THE EPI INDEX	**18th**	**75.46**	**28th**	**70.60**
ENVIRONMENTAL HEALTH	**20th**	**92.26**	**89th**	**63.87**
Air quality	29th	89.68	133rd	59.42

Household solid fuels	1st	100.00	62nd	63.46
PM2.5 Exposure	116th	80.58	162nd	50.78
PM2.5 Exceedance	115th	85.01	161st	62.69
Water and Sanitation	**14th**	**98.26**	**42nd**	**69.62**
Drinking water	20th	97.67	48th	65.12
Sanitation	18th	98.85	41st	74.12
Heavy metals	**13th**	**89.80**	**20th**	**87.21**
ECOSYSTEM VITALITY	**35th**	**64.25**	**3rd**	**75.08**
Biodiversity and Habitat	**78th**	**80.13**	**19th**	**94.31**
Marine protected areas	1st	100.00	-	-
Biome Protection	111th	66.62	1st	100.00
Species Protection Index	26th	99.82	1st	100.00
Representativeness Index	60th	57.50	43rd	68.25
Species Habitat Index	113th	77.72	79th	86.21
Forests	**32nd**	**35.79**	**82nd**	**17.09**
Fisheries	**131st**	**34.60**	**-**	**-**
Fish Stock Status	111th	46.72	-	-
Regional Marine Trophic Index	117th	22.48	-	-
Climate and Energy	**77th**	**52.55**	**9th**	**74.21**
CO2 Emissions Intensity (total)	130th	37.68	15th	78.21
CO2 Emissions Intensity	87th	36.52	36th	54.92
Methane Emissions Intensity	28th	85.54	40th	76.29
NO2 Emissions Intensity	25th	87.37	11th	98.23
Black Carbon Emissions Intensity	6th	98.70	26th	79.02
Air pollution	**5th**	**96.56**	**18th**	**79.51**
SO2 Emissions Intensity	3rd	98.72	43rd	70.22
NOX Emissions Intensity	9th	94.40	15th	88.81

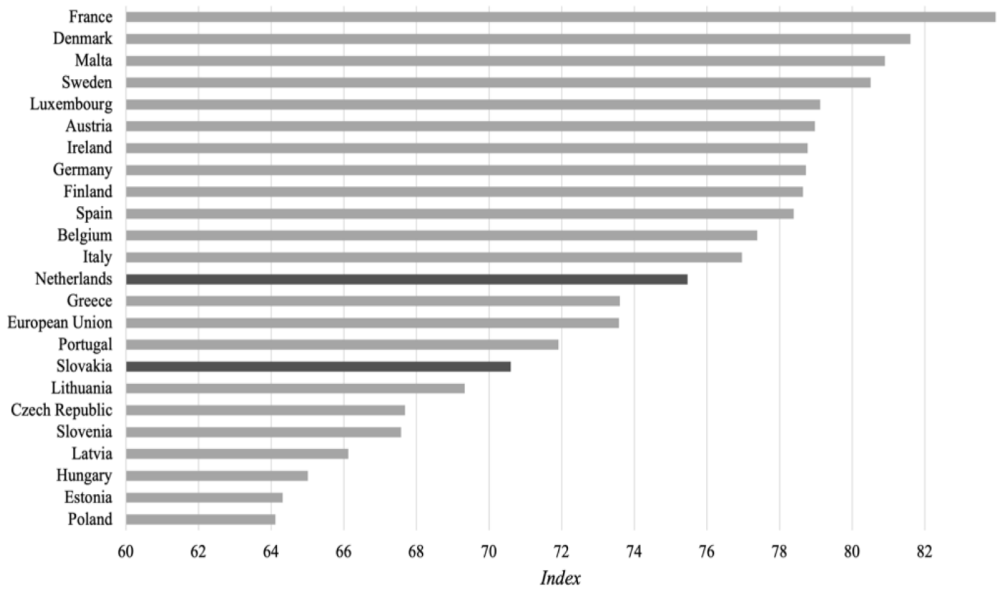
https://doi.org/10.53465/ER.2644-7185.2022.2.149-170

Water Resources	**3rd**	**99.90**	**46th**	**89.95**
Agriculture	**59th**	**35.39**	**13th**	**61.53**

Note: The missing data in the Biodiversity and Biotope: Marine Protection Areas and Fisheries sections show that Slovakia does not have access to the sea and the benefits of coastal states, and that the Netherlands does not have missing data as the country is washed by the North Sea from the Northwest.

Source: Own processing according to the EPI Index (2020a) and (2020b).

Figure 2: International Comparison of the Environmental Performance Index



Source: Own processing according to Environmental Performance Index (2018, 2020a, 2020b)

4 Hierarchical clustering procedures

The last presented analysis is a comparison of 3 classifications of cluster analysis on a set of 15 countries of the European Union. Our 2 examined variables were: Human Development Index (HDI) and Environmental Performance Index (EPI) as aggregated indicators, which we described in more detail in the introduction in the first chapter, from a methodological point of view in the third chapter and their application in the last chapter Results.

The characteristics of the raw data was considered in the selection of appropriate hierarchical clustering procedures. In the cluster analysis of our data, we used the statistical software SAS Enterprise Guide 4.2[5], which forms hierarchical clusters of observations containing the coordinates of the data, but also their distances. If the data set contains coordinates, the cluster analysis calculates the Euclidean distance of the objects before the clustering method is applied. The result of hierarchical agglomerative clustering is a graph displayed as a tree diagram - a "dendrogram", which can be displayed in the SAS system in 2 ways, vertically or horizontally. The main use of the dendrogram is to find the best way to assign objects to clusters, and the key to interpretation is to focus on the height at which the two different objects are connected.

Cluster analysis is an important tool for any study to identify possible intentions for convergence in living standards, education, GDP growth, life expectancy and environmental protection to measure overall progress in environmental sustainability.

One of Britain's professors of environmental economics, Paul Ekins, suggested in 2011 that there was a link between environmental performance and measures to improve environmental sustainability.

Ideally, these measures would include (Ekins, Anandarajah & Strachan, 2011):

1. development of better measurement and monitoring systems to improve the collection of environmental data, the so-called environmental data;

2. development of environmental policies focused on extremely weak areas;

3. communication of data and statistics at national level to international agencies such as the United Nations (UN);

4. the definition of sub-national metrics and targets for the improvement of environmental performance.

[5] Available on the SAS software website: (SAS, 2022) <https://www.sas.com/sk_sk/trials/software/covid19/form.htm>

4.1 Nearest Neighbor Method

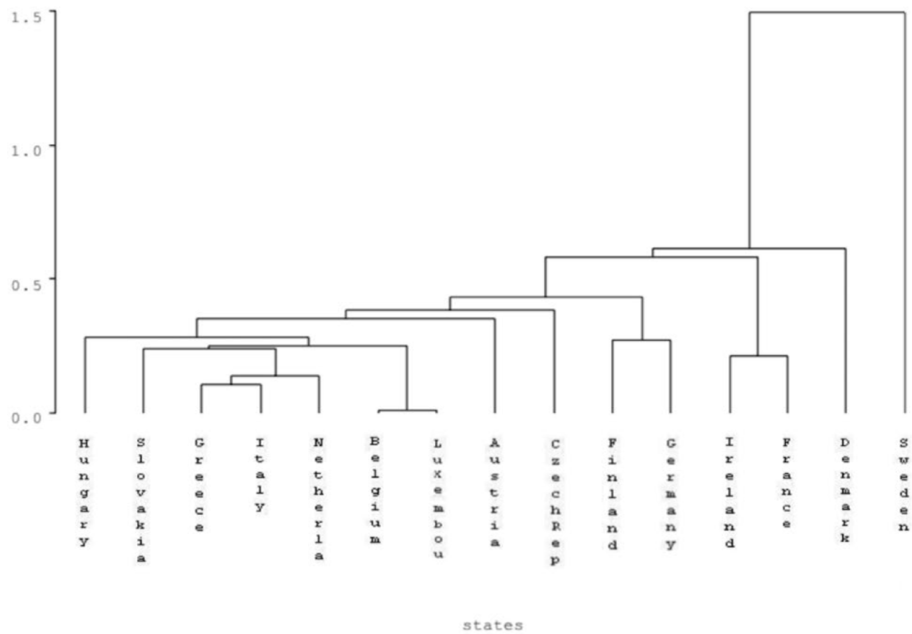
To compare the first cluster analysis classification, we used the *Nearest Neighbour Method* as the first of the hierarchical clustering methods. The principle of the nearest neighbour method is that the algorithm uses a minimum distance to measure the distance between clusters and 2 objects placed in a cluster are separated from each other by the shortest possible distance, gradually adding more clusters to the original objects by creating the 3rd nearest neighbour. After processing the classification using SAS, we constructed a dendrogram.

Table 4: Clusters according to the nearest neighbour method

Clusters	EU countries
1.	Sweden
2.	Hungary, Slovakia, Greece, Italy, Netherlands, Belgium, Luxembourg, Austria, Czech Republic, Finland, Germany, Ireland, France, Denmark

Source: Own processing according to data obtained from HDI and EPI index variables.

Figure 3: Cluster created according to the nearest neighbour method



Source: Own processing according to data obtained from HDI and EPI index variables.

According to the constructed dendrogram (Figure 3) and from Table 4 it follows visually and analytically that we divided the set of 15 countries into 2 clusters. If we take a closer look at the formed clusters, we can state that cluster 2, as a larger group, contains the predominance of 14 developed countries of the European Union. Countries such as Denmark, France, the Netherlands, Italy, Greece, Slovakia, and Hungary used the dendrogram to show a similar level of HDI and EPI indices. Cluster 1 is made up of only one EU country, Sweden, as significantly more advanced in terms of obtaining higher values of HDI and EPI indices.

4.2 Median Method

As the second method of cluster analysis for the comparison of European countries, we chose the *Median method,* which serves as a certain upgrade of the Centroid method. We have described the detailed principle of these methods in more detail in the previous chapters. The centroid method uses the distance between the centre of gravity of two clusters to evaluate the overall solution of the cluster, with the centre of gravity representing the centroid of a particular cluster. The distance between two clusters is calculated as the difference between the centres of gravity. The median method is based on the median, which follows from the name itself, and instead of calculating the average for each cluster to determine its centre of gravity, it calculates the mean distance between all pairs of observations or individuals in the clusters. After the data for this classification were processed, we built a dendrogram using SAS software.

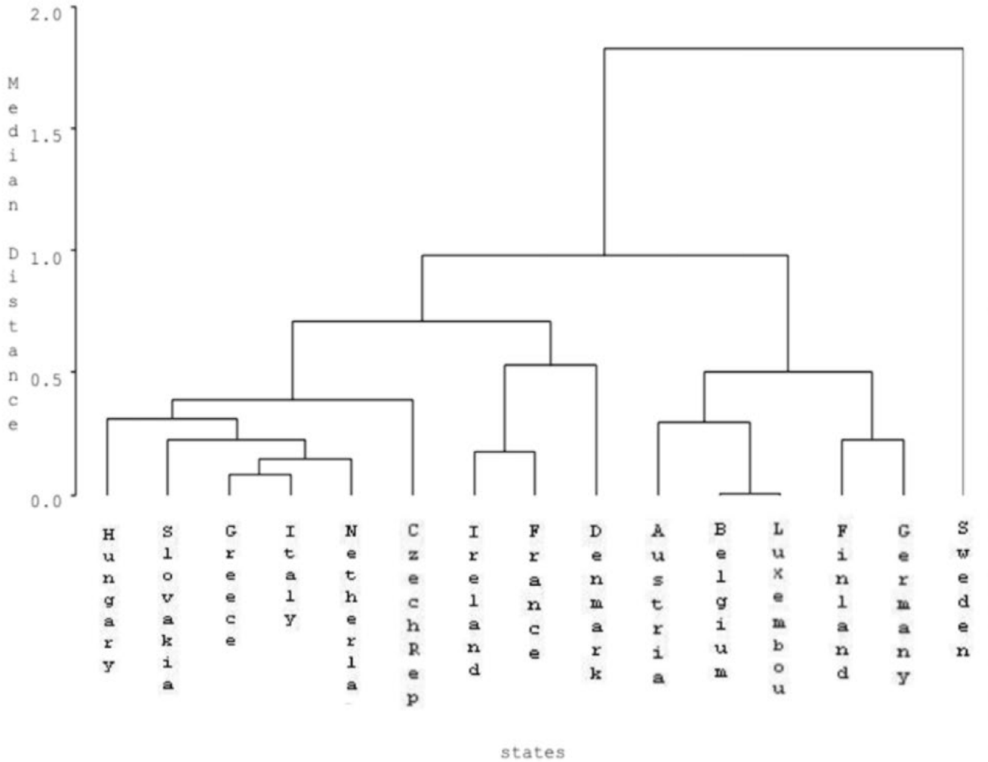
Table 5: Clusters according to the median method

Clusters	EU countries
1.	Sweden
2.	Hungary, Slovakia, Greece, Italy, Netherlands, Czech Republic, Ireland, France, Denmark, Austria, Belgium, Luxembourg, Finland, Germany

Source: Own processing according to data obtained from HDI and EPI index variables.

EKONOMICKÉ ROZHĽADY – ECONOMIC REVIEW, 2022, 51(2), 149 ─ 170
https://doi.org/10.53465/ER.2644-7185.2022.2.149-170

Figure 4: Cluster according to the median method



Source: Own processing according to data obtained from HDI and EPI index variables.

According to Table 5 and the dendrogram (Figure 4), we can observe a very similar situation as with the nearest neighbour method. We redistributed 15 countries into 2 main clusters. Cluster 2 contains again a set of 14 EU countries, whose monitored data of HDI and EPI indices are relatively similar. While Sweden belongs again to the 1st cluster and shows its strength over other countries, especially within the HDI and particularly in the dimension index called the *"Education index"*.

4.3 Ward's Method

As a final analysis, we present the most used method in marketing called *the Ward's Minimum Variance method.* Ward's method creates clusters that minimise variance in each cluster. For each cluster, the average for each variable is calculated and, in each cluster, the observations are compared to the average for each variable. The observations or clusters are combined in a way that the variance in the resulting cluster of solutions is minimised as much as possible. Following the summary of the data of our analysis, we prepared a table and constructed a dendrogram using SAS software.

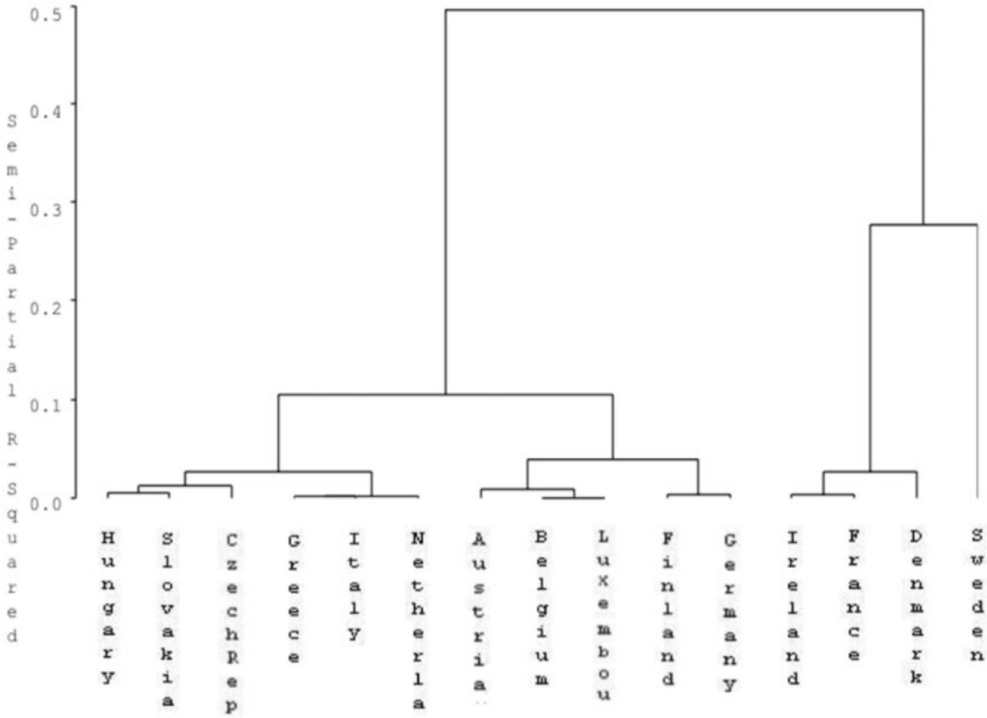
Table 6: Clusters according to the Ward's Minimum Variance method

CLUSTERS	EU countries
1.	Hungary, Slovakia, Czech Republic, Greece, Italy, the Netherlands, Austria, Belgium, Luxembourg, Finland, Germany
2.	Ireland, France, Denmark, Sweden

Source: Own processing according to data obtained from HDI and EPI index variables.

The illustrated dendrogram (Figure 5) illustrates the situation of 2 constructed clusters of countries, which can be very nicely distinguished from the cluster formed by Ward's Minimum Variance method. On the right side of the dendrogram we see cluster 2, which connects the 4 strongest countries in northern Europe. They are the world's richest economies with even income distribution, low unemployment, and highly developed institutionalisation, in terms of human data development index (HDI) and environmental performance index (EPI), what evokes a high level of standard in countries. From the opposite left side of the dendrogram, we can observe developed countries connected by one cluster with relatively similar values of the HDI and EPI indices. Although more significant differences can be seen mainly in countries such as Hungary (left side of the dendrogram) and Germany (closer to the Nordic countries of the dendrogram), where the differences are obvious and Hungary is trying to catch up, but it is not enough yet. Table 6 also clearly shows 2 clusters with a division of countries according to the achieved values of HDI and EPI indices.

Figure 5: Cluster according to the Ward's Minimum Variance method



Source: Own processing according to data obtained from HDI and EPI index variables.

When it comes to cluster properties, it is important to look at the values that countries indicate for the two indicators used for the analysis. In the case of cluster 1, things are clear: we have an economically strong country that seem to be operating under the control of the objectives of the European strategy and the appropriate values for an important environmental factor. In the case of clusters of the 2nd degree, we can observe interesting situations with all 3 analysed methods. In the case of Slovakia, we can see in the first nearest neighbour method how it reworked for the 2nd lowest position, which analyses that of all the countries studied, together with Hungary and Greece, it has the highest average of HDI and EPI indices. On the other hand, the Netherlands is approaching the average values of the indices to Belgium and Luxembourg. As defining features for the country in this grouping, we can say that they have an average employment rate between 70% and 81.1% (except Greece - 64% and Poland 64.60%) as well as high values for greenhouse gas emissions above 102 compared to 1990. These countries are the ones that need to make sustainable efforts to become knowledge-based economies. In the analysis of the median method, we get similar results as in the case of the first method, but the fundamental difference is the distance used between the centre of gravity of the two clusters to evaluate the overall solution of the cluster. However, it is more interesting in the last Ward's method, where the strongest EU countries (Sweden, Denmark, France, and Ireland) separated into a second cluster. The countries in cluster 1 seem to have interesting characteristics: greenhouse gas emissions are less than 71, compared to 1990 at 100, except for Belgium (92) and Sweden (91), and compared to cluster 2, the countries have a higher average of people at risk of poverty and lower average of primary consumption.

5 Conclusion

During the global development in various spheres of the environment, we have seen an increase in emissions in industry, waste, and emissions from transport. Consumption of raw materials is growing at the expense of environmental externalities, considering all stages of production of goods, from the extraction of raw materials to their waste management processes. Every day, a large amount of processed literature, reports from international and research centres, which point to the negative effects of environmental problems with impacts on climate change, are increasing. The changing climate is felt by society as a whole and affects almost every aspect of our lives. This may not only concern remote countries, but also EU countries such as Slovakia or the Netherlands.

Environmental behaviour is the result of many interrelated factors. Environmental inequalities in the environmental performance index are reported in all countries of the world, even the most developed ones. To create a sustainable and efficient green environment essential for human health, which would lead to the required ecosystem viability and environmental health, there must be cooperation between the environmental sector and other sectors in the country.

As a result of the use of more comprehensive approaches to measuring sustainable development, there is a direct increase in environmental awareness. Such approaches include the Human Development Index (HDI), which has been improving since the 1970s in both developing and developing countries. Improvements are reflected in the overall scale of education attained, life expectancy and rising living standards. Another approach is the Environmental

EKONOMICKÉ ROZHĽADY – ECONOMIC REVIEW, 2022, 51(2), 149 – 170
https://doi.org/10.53465/ER.2644-7185.2022.2.149-170

Performance Index (EPI), which allows countries to compare their social and environmental inequalities and clearly demonstrates that improvements and investments in the future are key to winning competitions in a sustainable environment: infrastructure (sanitary, water and electrical equipment), healthcare and education. Improving a country's partial competitiveness strengthens its long-term competitiveness. In our view, the development of sustainable competitiveness is also driven by social factors, which are playing an increasingly important role because of the constant growth of the average HDI.

Our paper has shown that geographical variations and typologies of environmental behaviour exist at the national level, and they can be identified. Based on research results on the impact of the Human Development Index (HDI) and the Environmental Performance Index (EPI) on environmental behaviour according to previous analysis results confirm a positive relationship.

In this context, there is extensive empirical research aimed to determine what influences environmental behaviour. For instance, one internationally renowned climate change expert, Karen O'Brien, examined in 2013 the impact of environmental health, which is one of the key determinants of health, air quality, water and sanitation, and its positive effects on environmental performance outcomes (O'Brien, Sygna & Wolf, 2013). One of the Australia's most famous politicians, Bob Brown, ranked among the most active environmentalists in 2014, also studied the factors that determine environmental behaviour. The author concludes that the geographical distribution of the degree of environmental behaviour requires different distributions, territorial differences and different types of development. Identifying and understanding geographical inequalities in environmental behaviour becomes an important mechanism required in any study focused on analysis and causation (Brown, 2014).

In this paper we made a comparison of 3 classifications of cluster analysis on a set of 15 EU countries using 2 examined variables of human development and environmental performance indices as aggregate indicators. During our multidimensional statistical classification, clusters were designed based on the HDI, EPI indices to evaluate the sustainable performance of EU members, as well as possible convergences between them at EU Member State level. The indicators used in the analysis form different groupings and most of the overlapping occurs in the groupings whose countries came first. This type of behaviour is typical of countries with strong economies, which record performance at all three socio-economic and environmental levels and pursue consistent development policies. Sweden and Denmark are the countries that appear in the first grouping in all analysed cases. Among the EU countries, Sweden appears most often in the leading grouping in all 3 analysed cases. The Czech Republic and Slovakia are ranked the best among the former communist countries and Luxembourg, Belgium, and the Netherlands as the third among the "Benelux" countries.

As part of the formulation of the benefits of the contribution, we would like to highlight the topicality of the issue, as the impact of climate change and environmental issues is currently one of the most discussed global issues. Care must be taken when analysing hierarchical clustering procedures, as cluster results are unable to provide an accurate picture of countries' sustainable performance assessments. To design a convergence model for EU Member States, a new indicator could be considered, which could help to design a comprehensive model for assessing the country's sustainable development and provide a framework for implementing best practice models in low-rated countries.

Acknowledgement

This paper is an outcome of VEGA 1/0646/20 “Diffusion and consequences of green innovations in imperfect competition markets” project solution.

REFERENCES

[1] Brown, B. (2014). *Optimism: Reflections on a Life of Action.* London: Hardie Grant Books.

[2] Dasgupta, S. (2016). A cost function for similarity-based hierarchical clustering. *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing,* 118 ─ 127. https://doi.org/10.1145/2897518.2897527

[3] Ekins, P., Anandarajah, G., & Strachan, N. (2011). Towards a Low-Carbon Economy: Pathways and Policy Requirements. *Climate Policy,* 11(2), 865 ─ 882. https://doi.org/10.1016/j.enpol.2012.08.027

[4] EPI. (2020a). Country Report - Netherlands. Available at: https://epi.yale.edu/epi-country-report/NLD

[5] EPI. (2020b). Country report - Slovakia. Available at: https://epi.yale.edu/epi-country-report/SVK

[6] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.* Waltham: Morgan Kaufmann.

[7] Jafarzadegan, J., Safi-Esfahani, F. & Beheshti, Z. (2019). Combining hierarchical clustering approaches using the PCA method. *Expert Systems with Applications,* 137, 1 ─ 10. https://doi.org/10.1016/j.eswa.2019.06.064

[8] Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika,* 3(32), 241 ─ 254. https://doi.org/10.1007/BF02289588

[9] Legendre, P., & Legendre, L. (2012). *Numerical Ecology.* Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-444-53868-0.50008-3

[10] Murtagh, F., & Contreas, P. (2011). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews,* 1(2), 86 ─ 97. https://doi.org/10.1002/widm.53

[11] Řezanková, H., Húsek, D., & Snášel, V. (2009). *Shluková analýza dat.* Pruhonice: Professional Publishing.

[12] O'Brien, K., Sygna, L., & Wolf, J. (2013). *A changing environment for human security: transformative approaches to research, policy and action.* London: Routledge. https://doi.org/10.1146/annurev-environ-032112-100655

[13] Tryon, R., C. (1939). *Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality.* Ann Arbor, Michigan: Edwards brother Inc.

[14] UNDP. (2020). Human development reports. Available at: http://hdr.undp.org/en/indicators/137506#

[15] Van Vark, G., N., & Howells, W. W. (1984). *Multivariate Statistical Methods in Physical Anthropology: A Review of Recent Advances and Current Developments.* Heidelberg: Springer. https://doi.org/10.1007/978-94-009-6357-3

[16] Yale Center for Environmental Law & Policy. (2020). Environmental Performance Index 2018. Available at: https://doi.org/10.1007/978-94-009-6357-3