International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-9, Issue-12; Dec, 2022
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.912.3

A clusters analysis of energy sector: Monitoring Brazilian matrix

Rodrigo Speckhahn Soares da Silva, Claudelino Martins Dias Junior

Postgraduate Program in Administration at Federal University of Santa Catarina (UFSC), Brazil

Received: 07 Nov 2022,
Receive in revised form: 01 Dec 2022,
Accepted: 06 Dec 2022,
Available online: 12 Dec 2022

***Keywords*— markets energy, Brazilian energy matrix, clusters, energy sector.**

***Abstract*— *The present study provides an overview of energy production in Brazil and in other 53 (fifty-three) countries, considering 18 (eighteen) variables of interest collected from the World Bank's database of years 1997, 2002, 2007 and 2012. With an essentially quantitative approach and used Principal Components Analysis (PCA), 10 (ten) variables were adopted that most contributed to the variance of the data. In parallel, through the clustering technique, the countries were grouped by common characteristics, observing that Brazil, even with privileged water resources, has walked in the opposite direction of the other countries, presenting an involution in some reference indicators. The main contribution of the study is the design of an analysis model capable of defining, preliminarily, a design of the Brazilian energy matrix from the end of the 90s to the year 2012, a period that includes the publication of the secondary data used.*

I. INTRODUCTION

Faced with a complex and dynamic panorama and, therefore, adaptations that cross almost all institutional environments, the apparent success of the planning activity needs to contemplate the use of useful and available information, combining the application of techniques that contain a quantitative approach and aided by conceptual structures to enable more qualified interpretations. Provost and Fawcett (2013b) state that understanding this process and its stages makes it more systematic and, consequently, less subject to errors and inconsistencies. In this sense, two techniques for effectively analyzing the data are explained: principal component analysis (PCA) and cluster analysis.

Principal Component Analysis (PCA) is used when large multivariate data sets are analyzed, as it allows for reducing the dimensionality (number of components or dimensions) of a data set. With it, it is possible to replace a set of original variables p with a smaller number q. This reduction is made by searching for linear combinations, a_1'x, a_2'x,…, a_q'x called principal components, which have a maximum variance for the data, and are not correlated with previous combinations a_k'xs. When solving this maximization problem, one finds vectors a_1, a_2, ..., a_q that are the eigenvectors of the S data covariance matrix, which correspond to the q largest eigenvalues. These eigenvalues provide the variance of their respective principal components (Jolliffe, 2005). The relationship between the sum of the first q eigenvalues and the sum of the variances of all the p original variables represent the proportion of the total variance in the original set of data, calculated by the first q principal components (Jolliffe, 2005).

Kaufman and Rousseuw (2005) define cluster analysis as the art of finding clusters in data. And Jain (2010) as a multivariate technique to evaluate natural clusters of a set of patterns, points, or objects. From an operational point of view, clustering is defined as a representation of n objects to find K clusters based on a measure of similarity such that the similarities between objects of the same cluster are high, while the similarities between different objects are low. Jain (2010) presents 3 (three) main objectives in the use of data cluster analysis: (i) underlying structure, that is, to obtain an overview of the data, formulate hypotheses, detect anomalies, and identify prominent characteristics; (ii) natural classification by identifying the degree of similarity between forms or organisms (phylogenetic relationship); and (iii) compression by organizing the data in order to summarize them through cluster prototypes.

Sun et al. (2017) and Jain (2010) claim that clustering plays an important role in business contexts, being widely used in market segmentation, product categorization, the analysis of customer profiles, behavioral patterns, and performance levels workflow types, and business models.

According to Ketchen Jr and Shook (1996), cluster analysis takes a sample of elements as an organization and groups them so that the statistical variance between the elements is minimized while the variance between the clusters is maximized. Studies of this nature were initially addressed in industrial and organizational economics literature with groups defined through sets with few variables.

This methodology is prevalent in disciplines involving multivariate data analysis and is widely used in the managerial context, as it allows summarizing a large amount of data from a simplified visualization of interpretation. Although the K-means algorithm is robust to handle relatively large volumes of data and fast requires that the number of clusters is specified and sensitive to outliers' presence (Lopes & Gosling, 2021).

II. THE DESCRIPTION AND DISCUSSIONS THE STUDY

Using both analysis techniques (PCA and Clustering) presented and having as reference the model proposed by da Silva et al (2022) and moreover searching to determine a representation for the Brazilian electrical matrix. The first part of propose of application to consist to used 18 variables with historical series from 1960 to 2020 preliminarily selected of 266 countries of World Bank's (2020) database, which were: ANE - Alternative and nuclear energy (% of total energy use); APD - Average precipitation in depth (mm per year); ATE - Access to electricity (% of population); CPI - Consumer price index (2010 = 100); EIN - Energy imports, net (% of energy use); EPC - Electric power consumption (kWh per capita); EPFNS - Electricity production from nuclear sources (% of total); EPFOS - Electricity production from oil sources (% of total); EPFRS - Electricity production from renewable sources, excluding hydroelectric (kWh); EPOGCS - Electricity production from oil, gas, and coal sources (% of total); EPTDL - Electric power transmission and distribution losses (% of output); EU - Energy use (kg of oil equivalent per capita); FFEC - Fossil fuel energy consumption (% of total); ICP - Inflation, consumer prices (annual %); IEWPP - Investment in energy with private participation (current US$); MEES - Methane emissions in energy sector (thousand metric tons of CO_2 equivalent); REC - Renewable energy consumption (% of total final energy consumption); RIFRT - Renewable internal freshwater resources, total (billion cubic meters).

All statistical analyses and procedures were performed in the computer environment R in version 4.0.5 with the aid of the packages "*tidyverse*" (Wickham et al., 2019), "*FactoMiner*", and "*factoextra*" (Kassambara & Mundt, 2020). Initially, Table 1 presents the descriptive statistics values (average, standard deviation, the minimum and maximum values, and the percentage of unavailable data (NA) for the sample.

Table 1: Descriptive statistics of the selected variables

Nº	Code	Variable	Average	Standard Deviation	Minimum	Maximum	% of NA
-	**Year**		-	-	1960	2020	-
1	**ANE**	Alternative and nuclear energy (% of total energy use)	6.35	9.93	0.00	71.54	54.66
2	**APD**	Average precipitation in depth (mm per year)	1219.00	811.00	18.10	3240.00	84.95
3	**ATE**	Access to electricity (% of population)	99.50	29.46	0.53	100.00	57.68
4	**CPI**	Consumer price index (2010 = 100)	68.50	310.14	0.00	20422.89	38.12
5	**EIN**	Energy imports, net (% of energy use)	-73.14	555.55	-17632.77	100.00	54.47
6	**EPC**	Electric power consumption (kWh per capita)	3220.37	4467.02	5.76	54799.17	54.76

Nº	Code	Variable	Average	Standard Deviation	Minimum	Maximum	% of NA
7	EPFNS	Electricity production from nuclear sources (% of total)	4.63	12.64	0.00	87.99	54.71
8	EPFOS	Electricity production from oil, gas, and coal sources (% of total)	23.99	30.24	0.00	100.00	53.90
9	EPFRS	Electricity production from oil sources (% of total)	4.88E+09	4.48E+10	0.00	1.65E+12	53.81
10	EPOGCS	Electricity production from renewable sources, excluding hydroelectric (kWh)	59.67	33.91	0.00	100.00	53.90
11	EPTDL	Electric power transmission and distribution losses (% of output)	12.66	9.03	0.00	88.02	55.18
12	EU	Energy use (kg of oil equivalent per capita)	2383.88	3034.15	9.55	40710.11	53.48
13	FFEC	Fossil fuel energy consumption (% of total)	66.59	30.47	0.00	100.00	54.97
14	ICP	Inflation, consumer prices (annual %)	23.79	332.40	-18.11	23773.13	38.09
15	IEWPP	Investment in energy with private participation (current US$)	1.09E+09	2.70E+09	5.00E+05	3.45E+10	92.63
16	MEES	Methane emissions in energy sector (thousand metric tons of CO2 equivalent)	24285.00	170798.80	0.00	3187680.00	23.46
17	REC	Renewable energy consumption (% of total final energy consumption)	30.55	30.71	0.00	98.34	54.19
18	RIFRT	Renewable internal freshwater resources, total (billion cubic meters)	530.60	3183.50	0.00	43392.16	84.72

The values obtained from the calculations for all available years are presented. It is observed that data are not available for all countries and years considered in the universe of the selected variables as observed in the column % of NA. Therefore, to perform the cluster analysis, it was necessary to exclude the not available (NA) values from the database the values, keeping in the database only the occurrences of countries with data available in all variables. Of the total, 111 occurrences were obtained. Table 2 presents the descriptive statistics for the new database.

Table 2: Descriptive statistics after excluding NA values

Nº	Variable	Average	Min	Max
-	Year	-	1992	2012
1	ANE	5.924	0.000	54.435
2	APD	1335.9	18.100	3240.0
3	ATE	82.450	6.197	100.000
4	CPI	79.37282	0.05003	129.47112
5	EIN	-9.577	-332.610	96.723
6	EPC	1464.26	50.33	6617.13
7	EPFNS	0.000	0.000	47.155
8	EPFOS	5.299	0.000	95.879
9	EPFRS	4.072e+09	0.000	1.325e+11

Nº	Variable	Average	Min	Max
10	EPOGCS	66.42	0.000	100.00
11	EPTDL	14.005	2.065	40.933
12	EU	1105.0	139.1	5167.0
13	FFEC	65.831	7.685	99.859
14	ICP	15.134	-10.068	951.962
15	IEWPP	1.540e+09	4.100e+06	3.035e+10
16	MEES	48821	170	677000
17	REC	32.8303	0.1804	92.7479
18	RIFRT	867.499	0.682	5661.000

To analyze principal components (PCA) and clusters, it is necessary to standardize the data according to the scale z (or z-score). Thus, the data were normalized (mean=0) with the unit of the new scale being a standard deviation, and the variance contained in each component was measured by eigenvalues, as shown in Table 3.

Table 3: Variance retained in each of the components

Component	Eigenvalues	Variance (%)	Cumulative Variance (%)
1	**4.890**	**27.167**	**27.167**
2	**2.449**	13.607	40.774
3	**1.974**	10.968	51.742
4	**1.418**	7.880	59.622
5	**1.360**	7.554	67.176
6	**1.117**	6.203	73.379
7	**0.984**	5.466	78.845
8	**0.817**	4.540	83.385
9	**0.684**	3.802	87.186
10	**0.601**	3.341	90.528
11	**0.473**	2.627	93.154
12	**0.403**	2.238	95.392
13	**0.294**	1.631	97.023
14	**0.199**	1.107	98.130
15	**0.176**	0.980	99.110
16	**0.119**	0.659	99.769
17	**0.023**	0.126	99.895
18	**0.019**	0.105	100.000

Components with eigenvalues greater than 1 (one) indicate that the variance of the component is greater than that, which would represent the variance of the original data. According to Battisti and Smolski (2019), this may be a criterion for deciding on the number of components to be used. The first 5 components correspond to almost 70% of the explanation of the variance of the data set (as highlighted in Table 3) and are therefore maintained for analysis purposes.

Besides Table 3 with the eigenvalues, it is possible to evaluate the correlation between the selected variables and each of the 5 components identified. Table 4 presents the measure of this correlation.

Table 4: Correlation between the selected variables and the 5 components obtained

Nº	Var	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
1	ANE	0.253	**0.582**	**-0.627**	0.150	0.054
2	APD	**-0.494**	0.275	**0.265**	-0.026	-0.239
3	ATE	**0.734**	-0.017	-0.111	0.039	**-0.340**
4	CPI	0.156	-0.208	-0.007	**0.739**	-0.258
5	EIN	-0.049	-0.057	-0.219	**0.326**	**0.723**
6	EPC	**0.904**	0.088	-0.077	-0.037	0.129
7	EPFNS	**0.513**	0.177	**-0.562**	0.006	0.267
8	EPFOGCS	0.218	**-0.807**	**0.340**	-0.096	0.133

Nº	Var	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
9	EPFOS	**-0.400**	**-0.445**	0.155	-0.110	**0.393**
10	EPFRS	0.246	0.217	**0.539**	**0.413**	0.136
11	EPTDL	-0.291	0.099	-0.193	-0.140	**-0.379**
12	EU	**0.858**	0.032	0.023	-0.192	0.114
13	FFEC	**0.779**	**-0.416**	0.052	-0.092	-0.222
14	ICP	-0.021	**0.345**	0.050	**-0.524**	0.147
15	IEWPP	0.238	**0.426**	**0.416**	**0.393**	0.001
16	MEES	**0.598**	0.182	**0.447**	-0.113	0.245
17	REC	**-0.843**	**0.345**	0.099	0.058	0.169
18	RIFRT	**0.338**	0.646	0.523	**-0.194**	**-0.004**

It is observed that 9 variables APD, ATE, EPFNS, EPFNS, EPFOS, REC, FFEC, EPC, EU, and MEES) showed a high correlation with Component 1. And ATE, EPFNS, EPC, EU, FFEC, and MEES (highlighted in green) are positively correlated (directly proportional), and APD, EPFNS, EPFOS, and REC (highlighted in red) are negatively correlated (inversely proportional). Thus Component 1 is directly related to access to electricity and non-renewable energy sources and is inversely proportional to energy consumption from renewable sources.

Similarly, Component 2 showed a positive correlation with the variables ANE, ICP, IEWPP, MEES, REC, RIFRT and a negative correlation with the variables EPFOGCS, EPFOS, and FFEC. Thus, component 2 is positively related to renewable energy sources and negatively related to non-renewable and fossil fuel-derived energies.

Concerning Component 3, this correlates positively with methane emission (MEES) and alternative energy sources other than hydroelectric (EPFRS). On the other hand, it correlates negatively with alternative nuclear energies (ANE).

In short, components 1 and 2 are responsible for explaining more than 40.8% of the total variance according to the data, especially Component 1 as shown in Figure 1.

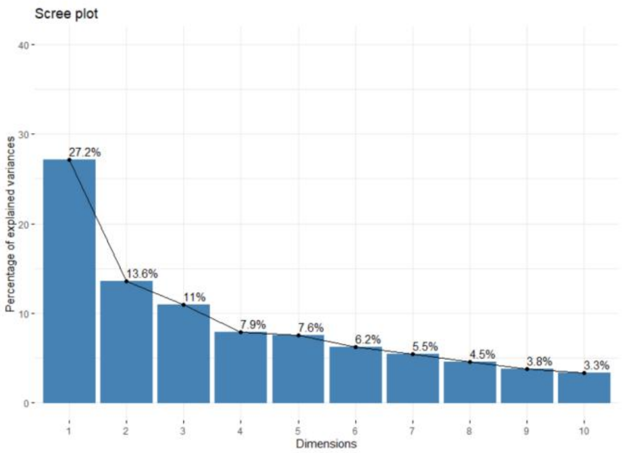

Fig 1: Percentage (%) of the variance of each component.

As shown in Table 4 and as shown in Figure 2, it can be seen which of the variables initially chosen for this study are the ones that most contribute to explaining the variance of Components 1 and 2.

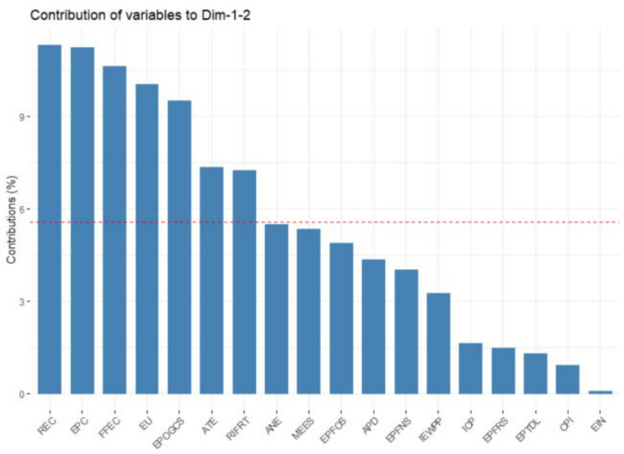

Fig. 2: Variables that most contribute to Components 1 and 2.

Thus, the cutoff was defined as the 10 variables (REC, EPC, FFEC, EU, EPOGCS, ATE, RIFRT, ANE, MEES, and EPFOS) that most contributed to the explanation of variance. The remaining 8 variables (APD, EPFNS, IEWPP, ICP, EPTDL, CPI, and EIN) were discarded because they contributed less than 1/5 of the total variance.

After excluding the NA, the countries' data were grouped in 1997, 2002, 2007, and 2012. These years and countries were selected, considering that they had all the data available.

Table 5: Countries grouped by year

Year	Countries	Nº
1997	Bangladesh, Bolivia, **Brazil**, Kazakhstan, Colombia, Philippines, Guatemala, India, Indonesia, Morocco, Mexico, Peru, Russia, Senegal, Tanzania, Vietnam, and Zambia	17
2002	Albania, Armenia, Azerbaijan, Bolivia, **Brazil**, Cameroon, Cambodia, China, Colombia, Philippines, Guatemala, India, Indonesia, Mexico, Nicaragua, Nigeria, Peru, Dominican Republic, Russia, Sri Lanka, Tajikistan, Tunisia, Ukraine, and Vietnam.	24
2007	Armenia, Bangladesh, Belarus, Bolivia, **Brazil**, Bulgaria, Cambodia, China, Colombia, Egypt, Philippines, Ghana, Georgia, Guatemala, India, Indonesia, Iran, Iraq, Jordan, Macedonia, Mexico, Moldova, Nicaragua, Nigeria, Pakistan, Peru, Romania, Russia, Sri Lanka, Ukraine, and Vietnam	31
2012	South Africa, Albania, Algeria, Bangladesh, Bosnia and Herzegovina, **Brazil**, Bulgaria, China, Colombia, Cote d'Ivoire, Costa Rica, Egypt, Philippines, Gabon, Ghana, Guatemala, India, Indonesia, Jordan, Malaysia, Morocco, Mexico, Mongolia, Nepal, Nicaragua, Pakistan, Peru, Kenya, Dominican Republic, Romania, Russia, Sri Lanka, Thailand, Turkey, Ukraine, Vietnam, and Zambia.	37

After defining the sets of countries to be considered for cluster analysis, it was determined how these countries were grouped according to their similarities and differences in each of the years considered. As for 1997, the countries were grouped into 2 clusters, as shown in the graph presented in Figure 3.

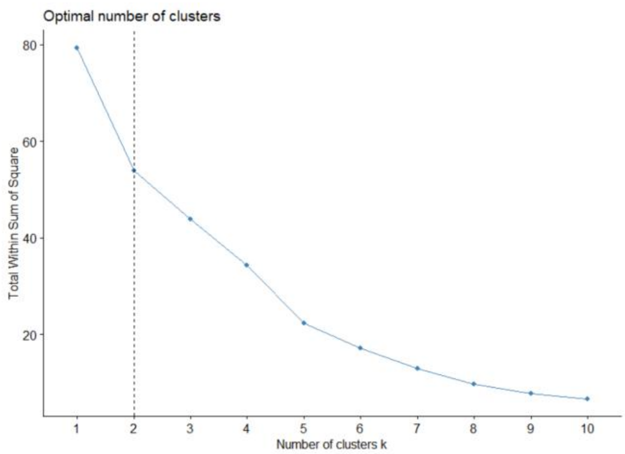

Fig. 3: Optimal number of clusters for 1997

According to Components 1 and 2, the countries were grouped as shown in Figure 4.

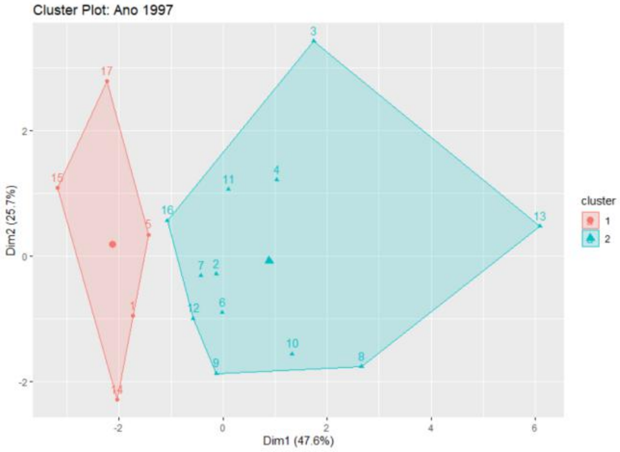

Fig 4: Clusters obtained for 1997

The countries belonging to Cluster 1 are Bangladesh, Bolivia, Guatemala, Senegal, Tanzania, and Zambia. They stood out negatively for having the worst levels of access to electricity (ATE) and positively for having the highest levels of renewable energy consumption. They also have the lowest levels of energy consumption from fossil fuels.

Cluster 2 includes Brazil (country # 3 in Figure 8), Colombia, Philippines, Indonesia, India, Kazakhstan, Morocco, Mexico, Peru, Russia, and Vietnam. Brazil stood out as the holder of the largest number of renewable waters (RIFRT) among all 17 countries.

It is also noteworthy in 1997 that, except for Russia, all presented energy use (EU) and electricity consumption (EPC) below the average of the countries present in this study. The full results for 1997 are presented in the Annexes section.

In 2002, the existence of 3 clusters was verified, as shown in Figure 5.

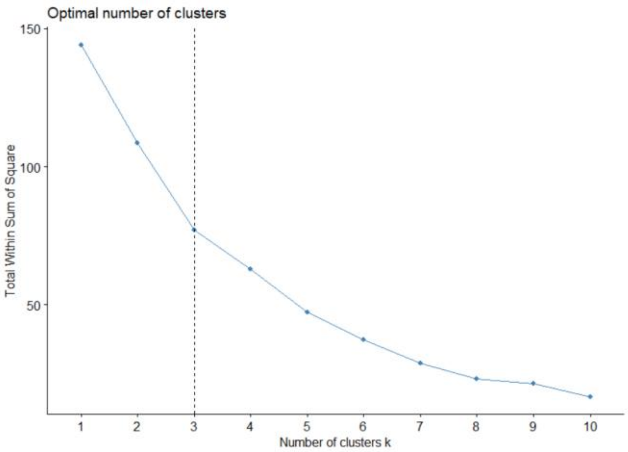

Fig. 5: Optimal number of clusters for 2002

The following countries belong to Cluster 1: Azerbaijan, Bolivia, China, Colombia, Dominican Republic, Indonesia, India, Mexico, Philippines, Russia, Tunisia, Ukraine, and Vietnam. These countries stand out from the others, mainly concerning the higher fossil fuel energy consumption (FFEC), electricity production from oil, gas, and coal (EPOGCS), and the percentage of their populations with access to electricity (ATE).

In 2002, once again, Russia stood out as the only country with per capita electricity consumption (EPC) and energy use (EU) above average among all the other countries in this study.

Cluster 2 includes Cameroon, Guatemala, Cambodia, Sri Lanka, Nigeria, and Nicaragua. The countries in this cluster, except for Cameroon, presented higher energy production values from oil, gas, and coal (EPOGCS), however, they presented the best numbers in relation to energy consumption from renewable sources. Also in this cluster, Cameroon, Cambodia, and Nigeria had the lowest fossil fuel consumption figures (FFEC).

Cluster 3 includes Albania, Armenia, Brazil (country # 5 in Figure 10), Peru, and Tajikistan. These countries presented the lowest results for energy production from oil, gas, and coal (EPOGCS). They also stood out in the use of alternative and nuclear-derived energies (ANE). The complete results for 2002 are also presented in the Annexes section.

The visualization of the clusters in the space defined by Components 1 and 2 for 2002 is in Figure 6.

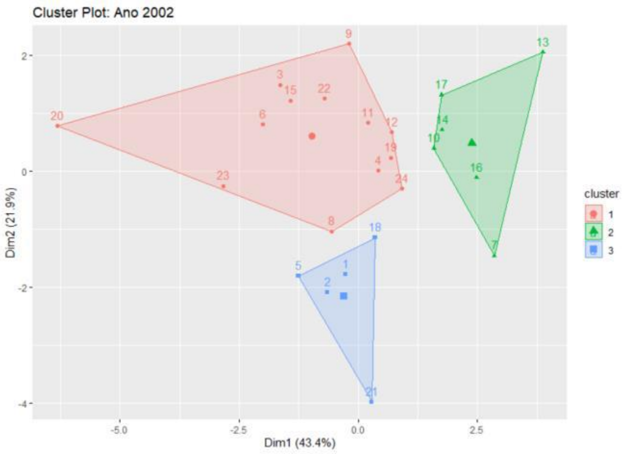

Fig. 6: Clusters obtained for 2002

In 2007, there are 2 (two) clusters of countries, as indicated in Figure 7.

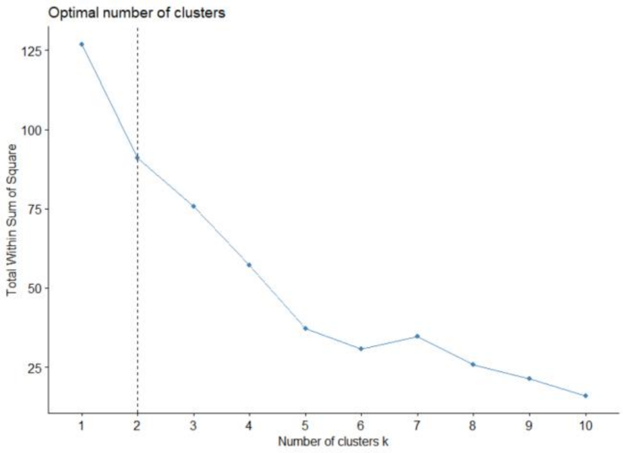

Fig 7: Optimal number of clusters for 2007

The countries belonging to Cluster 1 are Bangladesh, Brazil (country # 5 in Figure 12), Colombia, Georgia, Ghana, Guatemala, Indonesia, India, Cambodia, Sri Lanka, Nigeria, Nicaragua, Pakistan, Peru, Philippines, and Vietnam. In this cluster, it is noteworthy that all countries presented below-average indices for electricity consumption (EPC), energy use (EU), and fossil fuel consumption (FFEC). And values above the average for renewable energy consumption (REC). And to cluster 2 Armenia, Bulgaria, Belarus, Bolivia, China, Egypt, Iran, Iraq, Jordan, Moldova, Mexico, Macedonia, Romania, Russia, and Ukraine. This cluster of countries presented below-average values for renewable energy consumption (REC) and electricity production from oil sources (EPFOS). For the variables fossil fuel energy consumption (FFEC), energy production from oil, gas, and coal (EPOGCS), and access to electricity (ATE) above average. This year's full results are presented in the Annexes section, and clusters can be viewed in Figure 8.

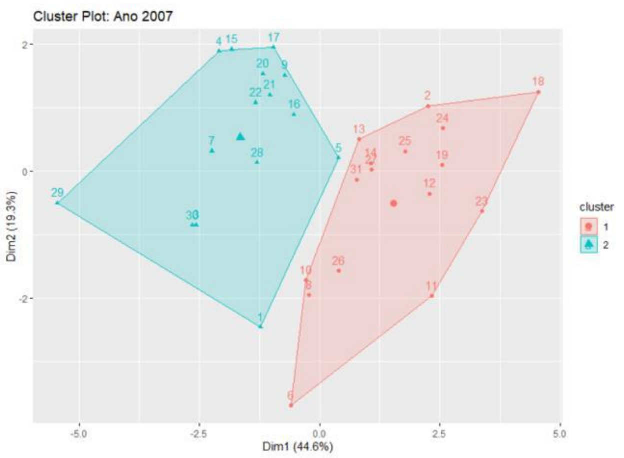

Fig. 8: Clusters obtained for 2007

For 2012, 2 clusters were also observed, as shown in Figure 9.

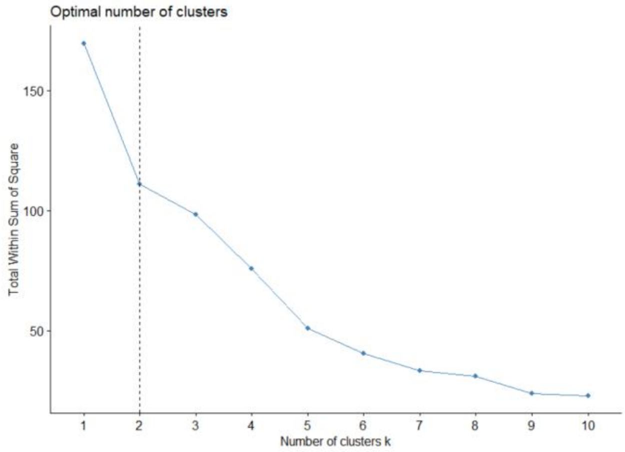

Fig. 9: Optimal number of clusters for 2012

The countries belonging to Cluster 1 are Bangladesh, Bulgaria, Bosnia and Herzegovina, China, the Dominican Republic, Algeria, Egypt, Indonesia, India, Jordan, Morocco, Mexico, Mongolia, Malaysia, Peru, Philippines, Romania, Russia, Thailand, Turkey, Ukraine, Vietnam, and South Africa. These countries showed low energy use generated from renewable sources (REC).

Cluster 2 includes Albania, Brazil (country # 5 in Figure 14), Cote d'Ivoire, Colombia, Costa Rica, Gabon, Ghana, Guatemala, Kenya, Sri Lanka, Nicaragua, Nepal, Pakistan, and Zambia. In this cluster of countries, it is observed that the variables per capita electricity consumption (EPC), fossil fuel energy consumption (FFEC), energy use (EU), and electricity generated from oil, gas, and coal (EPOGCS) below average and inverse behavior concerning energy consumption from renewable sources (REC).

The 2 (two) clusters showed similar behavior in relation to energy production from oil-related sources (EPFOS) and levels close to the average in relation to methane emissions (MEES), the exceptions to methane emissions were China (3.8215409) and Russia (3.7677935), which presented the highest emission levels.

The full results for 2012 are presented in the Appendix section, and the clusters can be viewed in Figure 10.

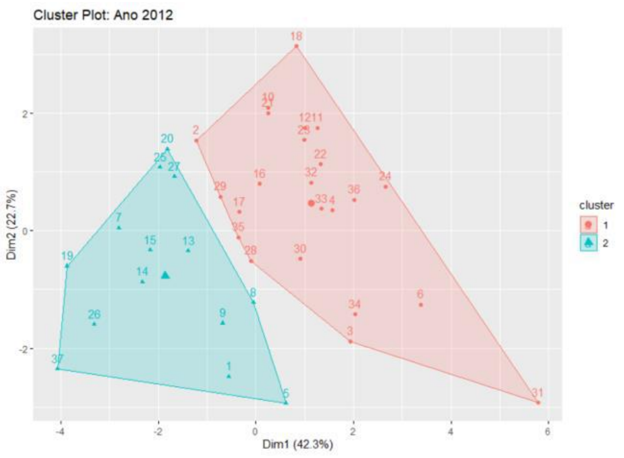

Fig 10: Clusters obtained for 2012

III. ANALYSIS OF RESULTS

The results of Brazil within the clusters over time were summarized in Chart 1. Its position was classified according to 3 (three) criteria: high (when grouped with the countries that presented the highest results), intermediate (grouped with the countries that presented intermediate values), and low (grouped with the countries that presented the lowest values) for each of the variables. Colors are a qualitative indication of the result. That is, green means that the numbers are relatively good, and red is considered bad. For example, the low level of methane emission was highlighted in green.

Table 6: Cluster in which Brazil is for each selected variable and in each year of analysis

Year	1997	2002	2007	2012
Alternative and Nuclear Energy (% of total energy used) (ANE)	High	High	Low	High
Access to electricity (% of population) (ATE)	High	Intermediate	Low	Low
Electricity Consumption (kWh per capita) (EPC)	Low	High	Low	Low
Electricity production from oil sources (% of total) (EPFOS)	Low	Low	High	Low
Electricity production from oil, gas, and coal sources (% of total) (EPOGCS)	Low, (2nd place with lower production)	Low, (3rd place with lower production)	Low, (2nd place with lower production)	Low (among the 5 lowest)
Energy use (kg of oil equivalent per capita) (EU)	Low	Low	Low	Low
Fossil fuel energy consumption (% of total) (FFEC)	High	Intermediate	Low	Low
Methane emission in the energy sector (thousand metric tons of CO2 equivalent) (MEES)	Low, except for Russia	Low, except for Russia and China	Low, except for Russia and China	Low, except for Russia and China
Renewable energy consumption (% of total final consumption) (REC)	Low	Intermediate	High	High
Renewable internal freshwater resources, total (billions of cubic meters) (RIFRT)	High, Brazil with the highest value	High, Brazil with the highest value	Equals, Brazil with the highest value	Equals, Brazil with the highest value

It is observed that Brazil was similar to countries that presented a reduction in fossil fuel energy consumption (FFEC) and those that increased renewable energy consumption (REC). As for electricity production from oil, gas, and coal sources (EPOGCS), it always remained among the 5 (five) countries with the lowest values. Regarding energy use (EU), it was always grouped with the countries that presented the lowest consumption levels. And concerning methane emission in the energy sector (MEES), it always remained at low levels. As for renewable internal freshwater resources (RIFRT), despite recurring water crises, it is verified that, historically, Brazil's position is privileged in relation to other countries in this regard.

In 2007, there was a change in the cluster in the use of alternative and nuclear energy (ANE), electricity production from oil sources (EPFOS), access to electricity (ATE), and electricity consumption (EPC). That is, it was grouped with countries that presented numbers considered poor. And this trend in the cluster remained for the ATE and EPC variables in 2012. In this context, more data would be appropriate to verify whether this trend has continued to the present day and how Brazil has advanced in the context of its energy matrix.

IV. CONCLUSIONS

As noted by Tolmasquim, Guerreiro, and Gorini (2007), the importance of the Brazilian energy sector cannot dispense with a continuous, systematized, and dynamic knowledge process in the face of the challenges of creating conditions for a rapid expansion of its supply that conditions the enhancement of productive investments on demand.

As suggestions for future studies aimed at deepening the results obtained, it is suggested to deepen the evaluation of the scales of the variables used and the application of other possible clustering techniques to obtain even more robust elements that support the rationale and plausibility of the proposed model as recommended by Lopes and Gosling (2021).

It is noteworthy that the objective of using quantitative variables to determine potential constituent elements of an energy matrix was achieved. Nevertheless, from the results obtained with clustering, it was possible to better understand the evolution of the status and repercussion of the Brazilian energy matrix compared to the other countries included in the analysis. It should be noted that the main limitation of the study focuses on the fact that the database used does not have complete data (NA) to obtain an even more complete and in-depth temporal analysis since it imposes limitations in the definition of the conditions found.

REFERENCES

[1] [1] Battisti, I. D. E., & Smolski, F. M. da S. (2019). *Software R: curso avançado*. https://smolski.github.io/livroavancado/

[2] [2] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

[3] [3] Jolliffe, I. (2005). Principal Component Analysis. Em B. S. Everitt & D. C. Howell (Orgs.), *Encyclopedia of Statistics in Behavioral Science* (p. bsa501). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013192.bsa501

[4] [4] Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*.

[5] [5] Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Wiley.

[6] [6] KETCHEN Jr., D. J., & Shook, C. L. (1996). THE APPLICATION OF CLUSTER ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH: AN ANALYSIS AND CRITIQUE. *Strategic Management Journal*, *17*(6), 441–458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

[7] [7] Lopes, H. E. G., & Gosling, M. de S. (2021). Cluster Analysis in Practice: Dealing with Outliers in Managerial Research. *Revista de Administração Contemporânea*, *25*(1), e200081. https://doi.org/10.1590/1982-7849rac2021200081

[8] [8] Silva, R. S. S. da, Dias Junior, C. M., Leite, A. L. da S., & Cario, S. A. F. (2022). A Representation Model of Brazilian Energy Matrix. *International Journal of Development Research*, *12*(10), 9. https://doi.org/10.37118/ijdr.25603.10.2022

[9] [9] Sun, L., Chen, G., Xiong, H., & Guo, C. (2017). Cluster Analysis in Data-Driven Management and Decisions. *Journal of Management Science and Engineering*, *2*(4), 227–251. https://doi.org/10.3724/SP.J.1383.204011

[10] [10] The World Bank. (2020). *The World Bank | Data* [Institucional]. Indicators. https://data.worldbank.org/indicator

[11] [11] Tolmasquim, M. T., Guerreiro, A., & Gorini, R. (2007). Matriz energética brasileira: Uma prospectiva. *Novos Estudos - CEBRAP*, *79*, 47–69. https://doi.org/10.1590/S0101-33002007000300003

[12] [12] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

APPENDIX

Table 7: Results of clusters of countries in 1997

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
1	Bangladesh	1997	0.9581	-0.7022	-0.3287	-0.7399	1.0105	-1.9983	-0.1337	-0.6095	-0.1293	-0.4814	1
2	Guatemala	1997	1.1317	-0.6666	-0.9299	-0.6109	-0.8223	-0.4198	-0.1324	-0.3775	-0.1274	0.2575	1
3	Senegal	1997	0.6552	-0.6946	-0.6406	-0.7126	1.0797	-1.6656	-0.1586	-0.6392	-0.1392	2.1572	1
4	Tanzania	1997	2.0252	-0.7084	-1.9331	-0.6666	-1.0372	-2.5483	-0.1403	-0.5379	-0.1100	-0.0540	1
5	Zambia	1997	1.8383	-0.5780	-1.7907	-0.5787	-1.7389	-2.1778	-0.1415	0.2084	-0.1302	-0.7803	1
6	Bolivia	1997	0.0720	-0.6340	0.3859	-0.6024	-0.1806	-0.4732	-0.0713	-0.3691	-0.1321	-0.7551	2
7	**Brazil**	**1997**	**0.4290**	**-0.3264**	**-0.3252**	**-0.4343**	**-1.5826**	**0.4101**	**1.6116**	**0.6640**	**-0.0470**	**-0.6869**	**2**
8	Colombia	1997	-0.1546	-0.5155	0.3039	-0.5460	-0.9233	0.4181	0.5071	0.1502	-0.0914	-0.7724	2
9	Indonesia	1997	0.5385	-0.6468	-0.1491	-0.5579	0.8844	-0.0330	0.4675	-0.5981	0.1775	-0.0780	2
10	India	1997	0.7144	-0.6369	-0.1581	-0.6547	0.6542	-0.9055	0.2875	-0.4531	0.2887	-0.6471	2
11	Kazakhstan	1997	-0.9403	-0.0105	1.0101	0.0632	0.8208	0.6109	-0.1465	-0.5327	-0.0682	-0.5521	2
12	Morocco	1997	-0.4242	-0.6199	0.6539	-0.6656	0.7374	-0.5828	-0.1576	-0.4878	-0.1259	0.4491	2
13	Mexico	1997	-0.5812	-0.3788	0.6586	-0.2920	0.4167	0.5111	-0.0382	-0.3038	0.0682	0.7854	2
14	Peru	1997	0.0388	-0.5872	0.0665	-0.6427	-1.0052	-0.4111	0.3488	0.2220	-0.1315	-0.0076	2
15	Philippines	1997	0.1059	-0.6174	-0.2910	-0.6205	0.2034	-0.3366	-0.0162	-0.3837	-0.1195	0.7950	2
16	Russian Federation	1997	-0.8727	0.3840	0.8363	0.5556	0.2308	0.6357	1.1878	-0.0010	2.4660	-0.5984	2
17	Vietnam	1997	0.9848	-0.6755	-0.7922	-0.6793	-0.6073	-0.0975	-0.0538	-0.3150	-0.0964	-0.4169	2

Table 8: Results of clusters of countries in 2002

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
1	Azerbaijan	2002	-0.9181	-0.2739	1.0083	-0.3195	0.8708	0.6357	-0.1641	-0.5220	-0.0528	0.2309	1
2	Bolivia	2002	-0.0737	-0.6247	0.4683	-0.6132	-0.4013	-0.5863	-0.0713	-0.2692	-0.1328	-0.7529	1
3	China	2002	-0.1162	-0.4534	0.4795	-0.4611	0.6250	0.5504	0.7169	-0.3955	1.8745	-0.6918	1
4	Colombia	2002	-0.0570	-0.5254	0.2743	-0.5825	-1.0665	0.4639	0.5071	0.2782	-0.0815	-0.7863	1
5	Dominican Republic	2002	-0.4089	-0.4097	0.7113	-0.5050	0.9277	0.2878	-0.1593	-0.5132	-0.1395	2.0803	1
6	Indonesia	2002	0.4594	-0.6275	-0.1326	-0.5351	0.7484	0.2148	0.4675	-0.5710	0.1056	-0.0214	1
7	India	2002	0.6597	-0.6290	-0.0742	-0.6468	0.7402	-0.6439	0.2875	-0.4390	0.3350	-0.6419	1
8	Mexico	2002	-0.6468	-0.3081	0.7198	-0.2799	0.6136	0.5643	-0.0382	-0.3558	0.0451	0.3437	1
9	Philippines	2002	0.0743	-0.6040	-0.3517	-0.6285	0.1383	-0.1512	-0.0162	-0.3321	-0.1146	-0.3639	1
10	Russian Federation	2002	-0.8821	0.4665	0.7930	0.6276	0.1686	0.6357	1.1878	0.1201	2.8874	-0.6913	1
11	Tunisia	2002	-0.5210	-0.4823	0.6722	-0.5248	1.1659	0.5611	-0.1654	-0.6305	-0.1304	-0.4592	1
12	Ukraine	2002	-0.9610	-0.0841	0.5808	0.1415	-0.2995	0.6357	-0.1494	0.9088	0.3066	-0.7777	1
13	Vietnam	2002	0.7133	-0.6364	-0.4699	-0.6505	-0.3097	0.2657	-0.0538	-0.2434	-0.0737	-0.3880	1
14	Cameroon	2002	1.7784	-0.6862	-1.6664	-0.6540	-1.6539	-1.2629	-0.0809	-0.3151	-0.1264	-0.6749	2
15	Guatemala	2002	1.0840	-0.6392	-0.7314	-0.5891	-0.0414	-0.0884	-0.1324	-0.4750	-0.1230	0.6109	2
16	Cambodia	2002	1.7086	-0.7097	-1.5742	-0.6815	1.0678	-2.1226	-0.1288	-0.6337	-0.1325	2.3770	2
17	Sri Lanka	2002	1.0275	-0.6531	-0.7220	-0.6439	0.0634	-0.0194	-0.1501	-0.4333	-0.1328	1.2508	2
18	Nigeria	2002	1.7597	-0.6975	-1.5368	-0.5467	0.0624	-1.2450	-0.0973	-0.5939	0.1623	-0.7932	2

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
19	Nicaragua	2002	0.7854	-0.6420	-0.6531	-0.6195	0.4645	-0.2501	-0.1176	-0.5315	-0.1374	1.7005	2
20	Albania	2002	0.1741	-0.3676	-0.1102	-0.5681	-1.5811	0.6357	-0.1582	1.1671	-0.1412	-0.5933	3
21	Armenia	2002	-0.7050	-0.4442	-0.1570	-0.5819	-0.9157	0.5678	-0.1645	3.0263	-0.1384	-0.7932	3
22	**Brazil**	**2002**	**0.4045**	**-0.3165**	**-0.3096**	**-0.4263**	**-1.4604**	**0.5221**	**1.6116**	**0.6947**	**-0.0312**	**-0.6646**	**3**
23	Peru	2002	0.3145	-0.5581	0.0268	-0.6413	-1.2556	-0.2338	0.3488	0.5645	-0.1305	-0.5109	3
24	Tajikistan	2002	1.0941	-0.2417	-1.0196	-0.6766	-1.7328	0.6018	-0.1467	4.8427	-0.1407	-0.7932	3

Table 9: Results of clusters of countries in 2007

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
1	Bangladesh	2007	0.5524	-0.6748	-0.0057	-0.7267	1.1181	-1.1803	-0.1337	-0.6172	-0.1256	-0.6180	1
2	**Brazil**	**2007**	**0.5466**	**-0.2357**	**-0.4571**	**-0.3775**	**-1.5015**	**0.5721**	**1.6116**	**0.7266**	**-0.0492**	**-0.6941**	**1**
3	Colombia	2007	-0.0041	-0.4996	0.2566	-0.5725	-1.2148	0.5023	0.5071	0.4026	-0.0676	-0.7841	1
4	Georgia	2007	0.1847	-0.3113	0.1346	-0.5004	-1.2236	0.6357	-0.1484	0.6466	-0.1336	-0.7836	1
5	Ghana	2007	0.7809	-0.6669	-0.6837	-0.6957	-1.7594	-0.8247	-0.1572	-0.2271	-0.1146	-0.7932	1
6	Guatemala	2007	1.0687	-0.5972	-0.7609	-0.5775	-0.2150	-0.0171	-0.1324	-0.3644	-0.1214	0.5122	1
7	Indonesia	2007	0.3082	-0.5987	-0.0658	-0.5259	0.8088	0.3336	0.4675	-0.5652	0.0851	0.0818	1
8	India	2007	0.4674	-0.5996	0.0425	-0.6258	0.6011	-0.3781	0.2875	-0.4073	0.4243	-0.6941	1
9	Cambodia	2007	1.2640	-0.6984	-1.1034	-0.7019	1.0549	-1.5677	-0.1288	-0.6277	-0.1311	2.3625	1
10	Sri Lanka	2007	0.9655	-0.6268	-0.6958	-0.6319	0.0079	-0.0431	-0.1501	-0.3717	-0.1310	1.1885	1
11	Nigeria	2007	1.8490	-	-	-	0.3902	-	-0.0973	-	0.2130	-0.7932	1

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
				0.6898	1.6083	0.5381		1.0570		0.5977			
12	Nicaragua	2007	0.7269	-0.6183	-0.5716	-0.6158	0.3345	-0.1601	-0.1176	-0.5196	-0.1371	1.5547	1
13	Pakistan	2007	0.4656	-0.6180	-0.1350	-0.6207	0.2096	-0.3592	-0.1494	-0.3158	-0.0618	0.2720	1
14	Peru	2007	0.1791	-0.5045	0.1236	-0.6189	-0.7778	0.0243	0.3488	0.4272	-0.1278	-0.6943	1
15	Philippines	2007	0.0214	-0.5907	-0.3057	-0.6432	0.2575	0.0268	-0.0162	-0.2993	-0.1138	-0.5076	1
16	Vietnam	2007	0.3773	-0.5579	-0.1267	-0.6102	0.1722	0.4066	-0.0538	-0.2377	-0.0420	-0.6621	1
17	Armenia	2007	-0.7663	-0.3240	0.1301	-0.4651	-1.0149	0.6096	-0.1645	2.1647	-0.1373	-0.7932	2
18	Bulgaria	2007	-0.7069	0.2917	0.3653	0.0930	-0.0173	0.6357	-0.1601	1.3774	-0.1340	-0.7494	2
19	Belarus	2007	-0.7579	0.0388	0.8451	0.1760	1.1839	0.6357	-0.1560	-0.6382	-0.1339	-0.7674	2
20	Bolivia	2007	-0.3240	-0.6003	0.5073	-0.6030	-0.0403	-0.0378	-0.0713	-0.3119	-0.1283	-0.7349	2
21	China	2007	-0.5101	-0.2002	0.6832	-0.2484	0.6834	0.5920	0.7169	-0.3447	2.6032	-0.7649	2
22	Egypt	2007	-0.8047	-0.4020	0.9665	-0.4926	0.8042	0.5850	-0.1664	-0.4660	-0.0610	-0.3812	2
23	Iran	2007	-0.9523	-0.2044	1.0602	0.0965	0.9272	0.5936	-0.1263	-0.5769	0.3517	-0.1360	2
24	Iraq	2007	-0.9294	-0.5459	1.0035	-0.4860	-0.1140	0.5712	-0.1556	-0.5107	-0.1351	-0.2351	2
25	Jordan	2007	-0.9188	-0.3214	1.0443	-0.4059	1.1725	0.5950	-0.1665	-0.4922	-0.1405	-0.2304	2
26	Moldova	2007	-0.8502	-0.1646	0.7537	-0.4006	1.0011	0.6357	-0.1662	-0.5343	-0.1376	-0.7904	2
27	Mexico	2007	-0.6867	-0.2626	0.7322	-0.2377	0.6580	0.5716	-0.0382	-0.3540	0.0867	-0.1236	2
28	North Macedonia	2007	-0.4834	0.0985	0.6061	-0.2978	0.7310	0.6357	-0.1650	-0.3093	-0.1396	-0.5494	2
29	Romania	2007	-	-	0.5444	-	0.0555	0.6357	-0.1534	0.1055	-0.0583	-0.7344	2

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
			0.4022	0.1546		0.1581							
30	Russian Federation	2007	-0.8752	0.6863	0.7863	0.7666	0.1990	0.6357	1.1878	0.1486	3.6315	-0.7370	2
31	Ukraine	2007	-0.9170	0.0691	0.4855	0.2018	-0.3606	0.6289	-0.1494	1.1587	0.2354	-0.7824	2

Table 10: Results of clusters of countries in 2012

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
1	Bangladesh	2012	0.2633	-0.6575	0.2000	-0.7132	1.1421	-0.5383	-0.1337	-0.6225	-0.1235	-0.3853	1
2	Bulgaria	2012	-0.4866	0.3451	0.1837	0.0427	-0.1481	0.6357	-0.1601	1.8312	-0.1328	-0.7776	1
3	Bosnia and Herzegovina	2012	-0.4972	0.0579	0.8322	-0.1746	0.3067	0.6357	-0.1555	-0.2089	-0.1371	-0.7866	1
4	China	2012	-0.6190	0.0550	0.7164	-0.0772	0.5365	0.6345	0.7169	-0.2272	3.8215	-0.7859	1
5	Dominican Republic	2012	-0.4544	-0.3853	0.5983	-0.5169	0.7624	0.5633	-0.1593	-0.4005	-0.1391	0.8621	1
6	Algeria	2012	-0.9888	-0.4438	1.0918	-0.3804	1.1573	0.5938	-0.1631	-0.6309	0.0336	-0.7322	1
7	Egypt	2012	-0.8224	-0.3436	0.9920	-0.4872	0.9275	0.6255	-0.1664	-0.5010	-0.0566	-0.2919	1
8	Indonesia	2012	-0.0574	-0.5567	-0.0409	-0.5045	0.8577	0.5000	0.4675	-0.5597	0.0653	-0.2986	1
9	India	2012	0.2865	-0.5590	0.1913	-0.5882	0.6375	-0.0465	0.2875	-0.3849	0.4589	-0.7258	1
10	Jordan	2012	-0.8988	-0.2916	1.0085	-0.4659	1.1769	0.6188	-0.1665	-0.4531	-0.1403	1.8851	1
11	Morocco	2012	-0.6387	-0.5246	0.7389	-0.6006	0.9348	0.4371	-0.1576	-0.5297	-0.1112	0.0437	1
12	Mexico	2012	-0.7028	-0.2164	0.8029	-0.2469	0.6960	0.6056	-0.0382	-0.3595	0.0695	-0.1886	1
13	Mongolia	2012	-0.898	-0.365	0.7739	-0.228	1.1893	0.0777	-0.1557	-0.639	-0.1345	-0.6195	1

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
			0	4		8				2			
14	Malaysia	2012	-0.9126	0.2503	0.9895	0.0971	0.9714	0.6289	0.0155	-0.5480	-0.0162	-0.6452	1
15	Peru	2012	0.0081	-0.4438	0.3304	-0.5573	-0.4398	0.3336	0.3488	0.1839	-0.1253	-0.7152	1
16	Philippines	2012	-0.0587	-0.5722	-0.2332	-0.6390	0.3498	0.1972	-0.0162	-0.3044	-0.1163	-0.6003	1
17	Romania	2012	-0.2930	-0.1379	0.3228	-0.2116	-0.1348	0.6357	-0.1534	0.4988	-0.0689	-0.7509	1
18	Russian Federation	2012	-0.8890	0.7604	0.7967	0.9173	0.2311	0.6357	1.1878	0.1302	3.7678	-0.7065	1
19	Thailand	2012	-0.2360	-0.1739	0.4454	-0.1726	0.9419	0.6055	-0.0962	-0.5275	-0.0271	-0.7451	1
20	Turkey	2012	-0.5705	-0.1010	0.7489	-0.2637	0.3848	0.6357	-0.0954	-0.0625	-0.0672	-0.7705	1
21	Ukraine	2012	-0.9016	0.0941	0.4157	0.0999	-0.3176	0.6312	-0.1494	1.3602	0.1686	-0.7843	1
22	Vietnam	2012	0.2459	-0.4509	0.0127	-0.5657	-0.1382	0.5641	-0.0538	0.0459	-0.0191	-0.7827	1
23	South Africa	2012	-0.6493	0.2565	0.6749	0.0833	1.0332	0.1368	-0.1526	-0.4021	-0.1119	-0.7907	1
24	Albania	2012	0.3093	-0.2467	-0.2900	-0.5590	-1.7594	0.6323	-0.1582	1.9207	-0.1411	-0.7932	2
25	**Brazil**	**2012**	**0.4234**	**-0.1609**	**-0.3295**	**-0.3197**	**-1.3289**	**0.6194**	**1.6116**	**0.6348**	**-0.0442**	**-0.6760**	**2**
26	Cote d'Ivoire	2012	1.4401	-0.6684	-1.3753	-0.5925	0.4097	-0.8646	-0.1425	-0.5569	-0.1353	-0.5915	2
27	Colombia	2012	-0.0536	-0.4590	0.2940	-0.5600	-1.1564	0.5350	0.5071	0.4279	-0.0493	-0.7750	2
28	Costa Rica	2012	0.2447	-0.2860	-0.6138	-0.4530	-1.5188	0.6189	-0.1312	0.8479	-0.1399	-0.5234	2
29	Gabon	2012	1.4626	-0.4996	-1.1780	-0.1026	-0.0643	0.2725	-0.1152	-0.4754	-0.1406	-0.3394	2

Nº	Country	Year	REC	EPC	FFEC	EU	EPOGCS	ATE	RIFRT	ANE	MEES	EPFOS	Cluster
30	Ghana	2012	0.5207	-0.6443	-0.4586	-0.6772	-1.4030	-0.8405	-0.1572	-0.0353	-0.0972	-0.7932	2
31	Guatemala	2012	1.2013	-0.5977	-1.1720	-0.5388	-0.7828	0.1839	-0.1324	-0.3366	-0.1202	-0.1331	2
32	Kenya	2012	1.5682	-0.6872	-1.6236	-0.6358	-1.0156	-1.4700	-0.1602	-0.4329	-0.0885	0.0408	2
33	Sri Lanka	2012	0.9816	-0.6035	-0.6326	-0.6041	0.3271	0.1945	-0.1501	-0.4063	-0.1295	1.1563	2
34	Nicaragua	2012	0.7399	-0.5935	-0.7648	-0.6046	-0.0733	-0.0585	-0.1176	-0.3864	-0.1370	1.0975	2
35	Nepal	2012	1.7632	-0.6939	-1.6867	-0.6623	-1.7436	-0.2220	-0.1044	-0.2053	-0.0941	-0.7755	2
36	Pakistan	2012	0.5422	-0.6251	-0.2331	-0.6332	0.1338	-0.3493	-0.1494	-0.2629	-0.0531	0.3948	2
37	Zambia	2012	1.8220	-0.5568	-1.8345	-0.5779	-1.7551	-1.8306	-0.1415	0.3395	-0.1261	-0.7884	2