

Analysis of International Eco-Economic Systems Using Data Mining Methods

Kharitonova A.E. *, Ulyankin A.E.

Russian State Agrarian University - Moscow Timiryazev Agricultural Academy, Moscow 127550, Russia

*Corresponding author. Email: kharitonova.a.e@rgau-msha.ru

ABSTRACT

The environmental pollution problems in the world have become urgent and are noted for sustainable development. The article analyzes 173 world countries in terms of ecological and economic systems and their development level, reflecting the relationship of the natural environment, economic and social human activity. A cluster analysis of countries by the environmental stress level was performed to determine the relationship of the demographic situation in the country, the level of medicine, education and economic development. The data mining methods were used to classify countries into groups by environmental indicators using sets of indicators, characterizing demography, medicine, education, and economic development. The developed methodological approach to the analysis is automated using R language.

Keywords: *natural environment, eco-economic system, environmental compliance, data mining, cluster analysis, environmental compliance*

1. INTRODUCTION

Throughout the existence of all mankind, the relationship of the natural environment state and the economy has always been an urgent issue; at present, however, this issue has become even more complex, namely, economic development progress on account of regression of the natural environment (NE) condition. Recently, statistical research of eco-economic systems is becoming more popular. The interest in the global environmental situation is called the "environmental boom". Today, economic activity has reached striking levels to be considered as a geological and climate forcing factor that can essentially change human life conditions [0].

In the whole world, forest areas are decreasing at an alarming rate every year, many species are still in danger of extinction. Besides, a spike in the number of crimes against wildlife, land-use changes including deforestation, and encroachment into animal and plant habitats provide conditions for new infectious diseases, including COVID-19, threatening public health and the global economy. More than 2 billion hectares, being one-fifth of the entire earth's land, are degrading. Land degradation has a negative effect on 3.2 billion people, resulting in the extinction of species and accelerating climate change. To overcome this situation, by 2020, 123 countries have set voluntary target values to halt the land degradation process (one of the global targets of the 2030 Agenda), in 60 countries these values have been officially approved. By changing the tendency to land degradation, it is possible not only to improve the well-being of billions of people and biodiversity, but also contribute to climate crisis resolution through measures for restoration, protection, and sustainable use of land [0].

But despite all efforts, climatic changes are still underway. Thus, 2019 was the second warmest year on record, and 2010–2019 decade was the warmest one. The global temperature increase by the end of the century may reach 3.2 °C. To keep the maximum gain rate, envisaged by Paris Agreement within the range of 1.5 °C - 2 °C, starting from 2020, greenhouse gas emissions shall be reduced by 7.6 percent every year. However, despite the sharp decline in human activity, resulting from COVID-19 crisis, the predicted 6% reduction in emissions for 2020 is below this index and emissions are expected to increase after removal of limitations [0].

Preservation of the natural environment and socio-economic development have always been in the focus of researchers. In developing countries, contradictions and conflicts between environmental protection and economic benefits are especially notable [0]. Over the past decades, processes of the natural environment oppression and economic activity growth resulted in a qualitative change of the entire global environmental system. These processes resulted in appearance of a new statistical object - eco-economic systems.

In our opinion, an eco-economic system (EES) is a NE, affected by the human economic and social activities, including interaction elements, as well as the resulting consequences. At the same time, eco-economic processes cover economic processes, affecting NE, as well as processes, arising in NE and are reflected in economic activity [0].

EES can be analyzed at the global (international), state, regional or district levels.

For EES analysis, the following can be considered as observation units:

- individual NE elements: for example, land plots, which can be used to study the qualitative chemical composition of soils as part of capability, etc.;
- enterprises or groups of enterprises, for example, by correlation emissions of harmful substances into the atmosphere or water resources with costs for NE protection;
- regional structures (cities, municipal districts, regions, districts, etc.), which can be used to study the effect of emissions into the atmosphere on human health, considering economic development of a region;
- different countries in terms of the national development level on EES and public health condition and quality [0].

Analytical tools for the analysis of eco-economic systems do not fully cover the interaction of environmental and economic factors, calling for improvement of EES study methods and techniques, as well as new automated analytic approaches.

Advances in sustainable consumption and production stimulate the achievement of sustainable development goals. During the period of 2017 and 2019, 79 countries and the European Union informed of the adoption of at least one national policy document to support the implementation of the 10-year strategy for sustainable consumption and production. In many countries, optimization models of the economy and environment interaction are under development. To do this, an eco-economic system simulation model was developed, it made it possible to make an ecological map of the Republic of Armenia [0]. This approach provides for possible scenarios of enterprise sustainable modernization as the main sources of emissions of harmful substances with the simultaneous development of the effective strategy for government control.

2. METHODOLOGY OF THE STUDY

A coordinated and unified set of indicators is being developed to ensure consistency and comparability of reporting data in the area of rational use of resources. However, the reports barely cover critical environmental, social and governance issues. There are also considerable differences between countries and regions. In East Asia, Latin America, the Caribbean, Europe and North America, companies present more thorough reports on key indicators to meet minimum requirements. In other regions, especially Africa, Central Asia and Oceania,

reported data are far from being complete [0]. The conflict between economic growth and environmental pollution has become a serious obstacle to the future development of China [0]. As a result, there is a need to differentiate the world countries for the development of eco-economic systems to create general provide common recommendations for similar countries. Similar recommendations were proposed for the distribution of the Russian Federation regions by the level of development to provide common recommendations for similar territories [0]. It is recommended to computerize statistics methodology using modern data processing tools for operational analysis [0].

The purpose of the analysis is to differentiate the world countries by the level of eco-economic systems and identify the relationship between sets of involved indicators.

To distinguish groups of countries according to the level of environmental development, the following relative indicators were used:

- x1 - eco-efficiency indicators;
- x2 - mortality rate attributed to environmental quality per 100,000 population;
- x3 - air quality index;
- x4 - drinking water quality index.

Data on 173 world countries were used for the analysis. The data were previously standardized. Five clusters were identified using the k-means method in Statistica package. Then the identified clusters were analyzed using indicators, characterizing the demographic situation, medicine, education level and economic development. Using data mining methods, identified groups were classified according to the environmental load using sets of indicators characterizing the demographic situation, medicine, education level and economic development. The following methods were used for classification: decision tree, random forest, support vector and discriminant mining methods.

3. RESULTS OF THE STUDY

According to the indicators, characterizing the environmental development of the country, 5 clusters of k-means methods were identified in the Statistica package. The distribution of countries by clusters is presented in Table 1.

Table 1 Distribution of countries by clusters

Cluster No.	Countries
1st cluster	Azerbaijan, Albania, Algeria, Armenia, Belize, Belarus, Bulgaria, Bolivia, Bosnia and Herzegovina, Vietnam, Gabon, Honduras, Georgia, Dominican Republic, Egypt, Iraq, Kazakhstan, Qatar, Kyrgyzstan, China, Lebanon, Macedonia, Maldives, Morocco, Moldova, Nicaragua, Oman, Paraguay, Peru, El Salvador, Samoa, Saudi Arabia, Saint Lucia, Serbia, Suriname, Thailand, Tonga, Tunisia, Turkmenistan, Turkey, Uzbekistan, Ukraine, CAR, Montenegro, Sri Lanka
2nd cluster	Antigua and Barbuda, Argentina, Bahamas, Barbados, Bahrain, Brazil, Brunei, Hungary, Venezuela, Jordan, Iran, Colombia, Costa Rica, Cuba, Kuwait, Latvia, Lithuania, Mauritius, Malaysia, Mexico, UAE, Panama, Poland, Russia, Romania, Seychelles, Slovakia, Slovenia, Trinidad and Tobago, Uruguay, Croatia, Czech Republic, Chile, Ecuador, Jamaica
3rd cluster	Angola, Bangladesh, Botswana, Bhutan, East Timor, Guyana, Guatemala, Djibouti, the Democratic Republic of the Congo, Zambia, Zimbabwe, Indonesia, Cape Verde, Cambodia, Kenya, Kiribati, Comoros, Congo, Madagascar, Malawi, Micronesia, Mozambique, Mongolia Myanmar, Namibia, Papua New Guinea, Rwanda, Eswatini, Senegal, Solomon Islands, Tajikistan, Tanzania, Uganda, Fiji, Philippines, Equatorial Guinea, Ethiopia, South Africa
4th cluster	Australia, Austria, Belgium, UK, Germany, Greece, Denmark, Israel, Ireland, Iceland, Spain, Italy, Canada, Cyprus, Luxembourg, Malta, Netherlands, New Zealand, Norway, Portugal, Singapore, USA, Finland, France, Switzerland, Sweden, Estonia, South Korea, Japan
5th cluster	Afghanistan, Benin, Burkina Faso, Burundi, Haiti, Gambia, Ghana, Guinea, Guinea-Bissau, India, Cameroon, Côte d'Ivoire, Laos, Lesotho, Liberia, Mauritania, Mali, Nepal, Niger, Nigeria, Pakistan, Sudan, Sierra Leone, Togo, Chad, Eritrea

According to the composition of clusters and the graph of means (Figure 1), it is obvious that the 4th cluster includes mainly developed countries, where eco-economic systems are characterized by the best indicators. These countries have the highest eco-efficiency, air quality, and drinking water quality indicators. At the same time, the mortality rate attributed to environmental quality per 100,000 population is lower than in other clusters. Therefore, it can be concluded that the 4th cluster is the best in terms of analyzed characteristics.

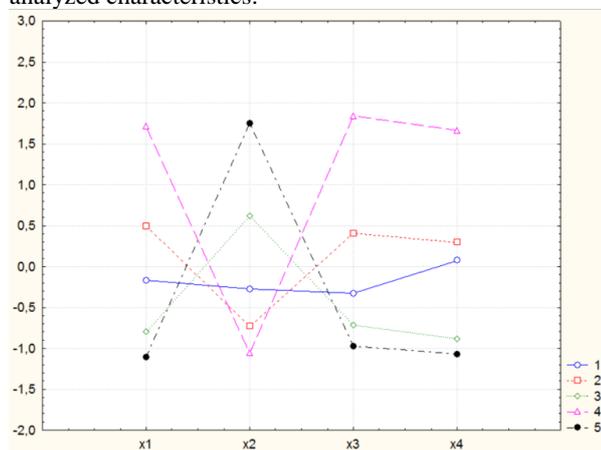


Figure 1 Graph of means by clusters

Cluster 2 ranks second in terms of environmental development level. In these countries mortality rate, attributed to environmental quality per 100,000, is low. Though, they have rather high indicators at the same time. The worst situation is in cluster 5. The countries with the highest mortality rate, attributed to environmental quality per 100,000, and the lowest indicators were included in this cluster. Mainly the countries of Asia and Africa were included in this cluster.

Clusters 1 and 3 occupy intermediate positions. At this, the environmental development level is higher in cluster 1 than in cluster 3 countries.

Let us characterize the identified clusters using indicators of the demographic situation, medicine, education level, and economic development.

Table 2 Characteristics of identified clusters

Indicator	Cluster No.					Average
	1	2	3	4	5	
Number of countries	45	35	38	29	26	173
Population density, people/km ²	78.10	24.01	64.32	31.28	135.34	55.74
Average area of 1 country, km ²	610,547	1,057,812	480,139	1,104,369	573,671	749,627
Education index	0.81	0.91	0.63	0.93	0.51	0.76
Human Development Index (HDI)	0.72	0.80	0.57	0.90	0.48	0.70
GDP per capita, USA	14,443	20,319	5,758	46,196	5,464	15,649
Urban residents in% of the total population	57.53	68.49	38.65	82.02	38.88	56.90
Unemployment rate, %	7.64	5.97	6.70	6.66	5.62	6.63
Adult obesity, %	21.86	25.87	10.92	21.76	5.40	17.78
Health expenditure in % to GDP	6.16	6.45	5.87	9.63	6.66	6.81
Birth rate per 1,000 people	17.57	14.58	28.15	11.00	33.20	20.54
Natural population growth per 1,000 people	10.30	7.19	19.57	2.23	23.20	12.29
Share of forests in the total country area	0.19	0.44	0.36	0.30	0.13	0.31
Share of agricultural land in the total country area	0.49	0.26	0.41	0.31	0.46	0.36

According to Table 2, it can be seen that in terms of the development level, the countries, included in cluster 4, surpass all the others. Cluster 4 is characterized by the highest Human Development Index (0.9), GDP per capita is almost 3 times higher than the average level by countries, the lowest population density and the highest share of urban residents in the total population. These countries are characterized by low birth rates, low natural population growth, the highest medical expenditure relative to the country's GDP, and a high level of adult obesity. Thus, it can be said that this cluster is characterized by the highest level of development of eco-economic systems. These countries address environmental problems and have a high level of economic and social development.

The second position in terms of the development level of eco-economic systems is occupied by the countries included in the 2nd cluster. They are also characterized by a fairly high level of economic development. GDP per capita is 30% higher than the average level for the analyzed countries, the Human Development Index was 0.8 (with an average level of 0.7). These countries have a low birth rate and natural population growth, a high share of forests in the total country area, and the lowest share of agricultural land. It should be noted that Russia belongs to this group, its indicators could have affected these two indicators.

The countries of cluster 1 are characterized by an average level of eco-economic development. For almost all indicators, the data correspond to the average level according to sampling of countries.

The countries, included in cluster 3, have a rather low level of eco-economic development. The countries of this group are characterized by low GDP per capita (below the average by 63%), a low Human Development Index (0.57), the lowest share of urban residents and the lowest health expenditure.

The lowest level of eco-economic development is observed in the countries of the 5th cluster. With a low environmental development a low economic development level is observed at the same time. The countries of this cluster have the lowest Human Development Index (only 0.48 with an average level of 0.7), the lowest GDP per capita (below the average level by 65%), low health expenditure, with the highest population density, birth rate and natural population growth.

According to the data, the sets of indicators are interrelated. It is possible to identify the relationship using data mining methods in the R environment. Table 3 presents the results of the predicted environmental development level of countries by sets of indicators, characterizing the demographic situation, medicine, education level, and economic development.

Table 3 Classification results of environmental development level, %

Classification method	Sets of indicators			
	demography	health care	education	economic development
Random forest	49.7	63.6	66.5	59.0
Support vector machine (SVM)	60.1	75.1	70.5	66.4
Discriminant mining	50.3	68.2	72.3	54.3
Decision tree	67.1	75.7	78.6	69.9

4. DISCUSSION OF RESULTS

In general, the classification results shall be considered as average. The best results were obtained using the decision tree algorithm. The indicators, characterizing the level of health care and education by country, have the greatest effect on the environmental development level. The decision tree algorithm provides statistically significant cluster classification, identified by the environmental load level by the indicators, characterizing the country education level by 78.6%. The demographic indicators have the least effect on the environmental load (the best classification accuracy is 67.1%).

To analyze and classify countries by the development level of eco-economic systems, an automated system was created to load data, divide countries into clusters according to the selected system of indicators and classify countries into groups according to the selected set of indicators.

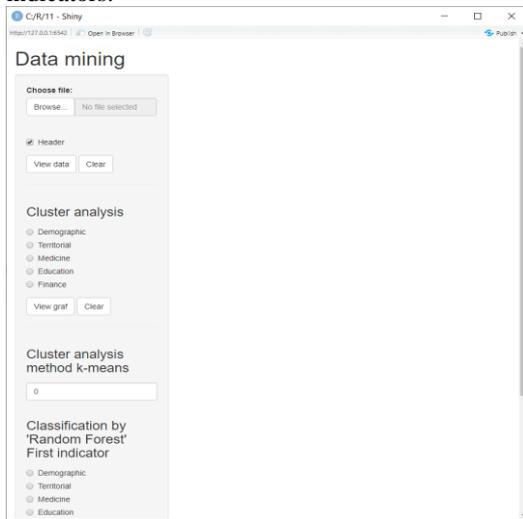


Figure 2 Application interface

The system was developed in the R environment using the Shiny library. This system will assist in the identification of relationships and dependencies between sets of indicators and quick information acquisition.

The solution of global environmental problems requires medium and long-term strategies for sustainable development and transformation in all countries of the world [0]. It is necessary to strive for the development of eco-economic systems at micro- and macrolevels.

5. CONCLUSIONS

As a result of the study, 173 world countries were analyzed by the level of eco-economic systems development, the automated system was developed to identify the relationships between sets of statistical indicators. In the future, this system can be improved with respect to main provisions of the Natural Resource and Economic Accounting System (NR EAS).

At present, all developed world countries are taking efforts to improve the macrostatistical study of eco-economic systems. One of the main activities of leading international economic organizations is the practical implementation of the main provisions of the Natural Resource and Economic Accounting System (NR EAS) based on the NR EAS Basic Scheme, adopted in 2012 as an international statistical standard [0]. In 2017, 69 countries implemented Natural Resource and Economic Accounting System programs, up by 28 % compared to 2014. At the same time, the formal acceptance of the Basic Scheme encouraged an additional and tangible incentive for relevant activities.

ACKNOWLEDGMENT

The article was prepared with the financial support of the Russian State Agrarian University-Moscow Agricultural Academy named after K.A. Timiryazev, the topic of the project is "Methodology of statistical research of ecological and economic systems of agriculture" No. 7.

REFERENCES

[1] Yu. Vertakova, V. Plotnikov, Problems of sustainable development worldwide and public policies for green economy, *Economic Annals-XXI*, 166(7–8) (2017) 4-10.

[2] Sustainable Development Goals Report 2020, UN 2020. https://unstats.un.org/sdgs/report/2020/The-Sustainable-Development-Goals-Report-2020_Russian.pdf

[3] J. Wu, Y. Guo, J. Zhou, Nexus between Ecological Conservation and Socio-Economic Development and its Dynamics: Insights from a Case in China, *Water* 2020, 12, 663

[4] A.D. Dumnov, L.I. Khoruzhy A.E. Kharitonova A.V. Ukolova, S.A. Skachkova, Eco-economic systems of Russian agriculture: statistical analysis, *J. of Reviews on Global Economics*, 8 (2019) 362-372

[5] A.D. Dumnov, A.E. Kharitonova, Statistical analysis and modeling of ecological and economic processes in agriculture, Moscow: Publishing house of the RSAU-Moscow Agricultural Academy, 160 p., 2016.

[6] System of Environmental-Economic Accounting: Central Framework/White cover publication, pre-edited text subject to official editing. European Commission, FAO, International Monetary Fund, OECD, United Nations, World Bank, 2012.

[7] L.A. Beklaryan, A.S. Akopov, A.L. Beklaryan, A.K. Saghatelian, Agent-based simulation modeling for regional ecological-economic systems. A case study of the Republic of Armenia, *Machine Learning and Data Analysis*, 2(1) (2016) 104–114. DOI: 10.21469/22233792.2.1.08

[8] L. Jing, Sh. Xing, W. Huaqing, L. Liwen, Trade-off between economic development and environmental governance in China: An analysis based on the effect of river chief system, *China Econ. Rev.* 60 (2020). DOI: 10.1016/j.chieco.2019.101403

[9] A.E. Kharitonova, Regional Differentiation by Indicators of Environmental- Economic State and Development of Agriculture, *Voprosy statistiki*, 25 (10) (2018) 37-46.

[10] A.E. Kharitonova, A.V. Tikhonova, M.V. Kagirova, A.M. Kozhemyakina Automation of Environmental and Economic Systems Research Using Data Mining, *Proc. of the Int. Conf. on Policies and Economics Measures for Agricultural Development* DOI: <https://doi.org/10.2991/aebmr.k.200729.067>

[11] L. Sievers, B. Breitschopf, M. Pfaff, A. Schaffer Macroeconomic impact of the German energy transition and its distribution by sectors and regions, *Ecological Economics*, (160) (2019) 191-204.