

Strategies development for the formation of resource-saving regional energy systems based on data mining methods

Evgeny Lisin^{1,*}, Galina Kurdiukova¹, and Irina Zameshaeva¹

¹National Research University "Moscow Power Engineering Institute", Department of Economics in Power Engineering and Industry, 111250 Moscow, Russia

Abstract. The Russian regions have been tasked with reducing the energy intensity of the country's gross product through the implementation of regional state programs for the formation of resource-saving energy systems. The required reduction in the GRP energy intensity can be achieved by different types of strategies for the development of the energy economy of the regions, the effectiveness of which largely depends on the production, technological and structural features of regional energy systems. In the work, using data mining tools, clustering and classification of regional power systems are carried out and the best resource saving strategies are proposed for each of the selected groups of power systems.

1 Introduction

A characteristic feature of the territorial organization of the energy complex of Russia is not the isolated location of its production facilities, but the work of their predominant part in the energy systems.

So in the electric power industry, power systems are a complex of large power plants of various types, interconnected by high-voltage power lines. Power systems contribute to the territorial dispersal of production and population, and can significantly reduce the required total capacity of power plants.

The natural desire of the regions for self-government orients the regional energy policy towards the development of the regional energy system along the path of self-sufficiency of their needs for energy resources and products (electricity and heat). At the same time, it is important to maintain close ties, technological and organizational unity of the regional energy system with the fuel supply and energy supply systems of the country.

The level of energy efficiency in the country and its regions is characterized by the energy intensity indicator of the gross domestic product. Russia is in last place in the world ranking of 44 countries in terms of energy consumption per unit of GDP, its energy intensity is 5 times higher than the best world values. The high energy intensity is due, in addition to climatic conditions, which are the cause of the long heating season (75% of the energy consumed by the population is heat), low efficiency of energy-intensive industries, as well as insufficient attention to the problems of energy conservation at the territorial level. In the regions of the country, there is a shortage of heat power, estimated at 20% of the total demand for heat energy, which leads to an overconsumption of fuel due to the use of other

energy sources by 2-2.5 times (when replacing heat with electricity, by 3.5-4 times). For comparison, the energy intensity of the GDP of Canada with a similar climate and developed industry is 1.8 times lower than the energy intensity of Russia [1, 2, 3].

The gradual transition to a post-industrial economy leads to a reduction in the share of energy-intensive industrial production and an increase in the share of household consumers in the structure of energy consumption. This is due to an increase in urban population and an increase in energy consumption per capita due to an increase in the number of household appliances that consume electricity. If in the case of an industrial consumer, on average in the energy system, it is possible to manage demand by stimulating the operation of an enterprise in several shifts, and thus to equalize the consumption of energy products in time, then for a household consumer this opportunity is absent, since the distribution of energy consumption by a person in time is due to his social and poorly regulated biological rhythms. This circumstance has a strong impact on the production efficiency of the power system.

The Russian regions have been tasked with reducing the energy intensity of the country's gross product by 1.5 times by 2035 through the implementation of regional state programs for the formation of resource-saving energy systems [4, 5]. The required reduction in the GRP energy intensity can be achieved by different types of strategies for the development of the energy economy of the regions, the effectiveness of which largely depends on the production, technological and structural features of regional energy systems. In the paper, using data mining tools, it is proposed to cluster and classify regional power systems in order to determine the best resource saving strategies for each of the selected groups of power systems.

* Corresponding author: lisinyam@mpei.ru

2 Background

The scientifically grounded direction of the development of the regional energy complex, ensuring the achievement of the goals of the energy policy of the region, as well as the mechanisms for its implementation are contained in the energy strategy of the region [6]. The basis for its development is the country's energy strategy, which considers regional aspects of the development of the energy complex for the next 20 years. Every 5 years the strategy is concretized [5, 7, 8].

The energy strategy provides for an individual approach to the development of fuel supply and energy supply to the country's regions. It is based on the volume and structure of demand for energy resources and electricity, estimated by making a forecast of the production of the gross regional product.

Table 1 presents summary data on the directions of regional development, grouped by federal districts, obtained on the basis of an analysis of the country's energy strategy until 2035. When forming the directions, the development scheme of the energy complex and the energy potential of the regions were taken into account.

Table 1. Development strategies for the energy complex at the regional level until 2035.

Federal district	Energy forecast of the region for 2035
Central Federal District	An energy-deficient region with a developed electricity and oil refining industry. Self-sufficiency in energy resources will increase to 21% due to the development of nuclear energy and the full implementation of energy conservation
Northwestern Federal District	An energy-surplus region that supplies energy resources and electricity to energy-deficient regions. Self-sufficiency in energy resources will increase to 200% due to an increase in energy production and the development of nuclear energy
Southern Federal District	Energy-deficient region, representing a transit and export hub. A moderate increase in energy demand and a significant increase in domestic production will increase the region's self-sufficiency to 60%
North Caucasian Federal District	An energy-deficient region, which is a transit hub. Uniform growth in energy consumption and production will lead to the preservation of self-sufficiency in the region at 80%
Volga Federal District	An energy-deficient region with declining production of its own energy resources and a developing processing industry. The level of self-sufficiency of the region will decrease to 55%
Ural Federal District	An energy-surplus region, which is the main supplier of energy carriers to energy-deficient regions and for export. The level of self-sufficiency in the region will remain at 120%
Siberian Federal District	An energy surplus region with a developed mining and processing industry (first in coal production, second in oil and gas). The level of self-sufficiency in the region will remain at 150%
Far Eastern Federal District	An energy surplus region with a developing mining industry. The level of self-sufficiency in the region will rise to 120%

The energy saving potential in the country as a whole is estimated at 40%. Figure 1 shows the contribution of the energy intensity of the gross regional product to the energy intensity of the country's gross product [9, 10].

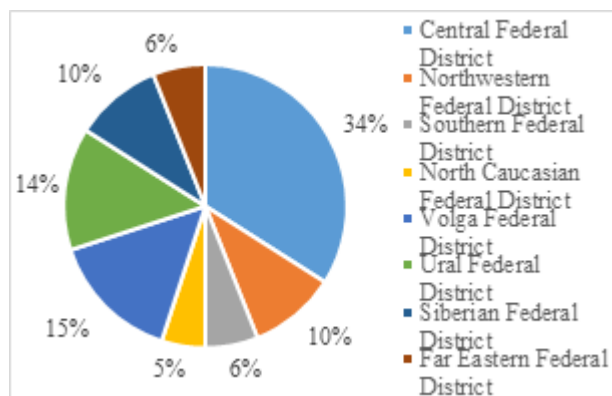


Fig. 1. Contribution of the energy intensity of the gross regional product to the energy intensity of the country's gross product by federal districts.

The required reduction in the energy intensity of the regions' GRP is achieved by the following types of strategies:

- reducing losses of energy resources and ensuring their energy conservation in various sectors of the region's economy (primarily, the energy sector),
- increase in GRP growth due to the organization of production with low energy intensity, in other words, the development of small business and services,
- decommissioning of old and inefficient production equipment and optimization of power system operating modes,
- assimilation of energy efficient production technologies by the regional energy complex.

The production, technological and structural differences and features of the power systems of the regions determine different priorities and scenarios for the implementation of the above strategies. Evaluation of their effectiveness requires a grouping of power systems in accordance with their general properties with further correlation of the selected strategies in accordance with the effect on energy saving achieved from their application.

3 Methods

One of the recognized approaches to the analysis of the energy intensity of the energy systems of the country's regions is the study of their groupings according to the specific consumption of energy resources and electricity [11, 12, 13]. The increased energy consumption per person characterizes the energy saturation of the region and the development of its energy system from the standpoint of production and technological potential and energy ties. With unit energy consumption up to 5 tce / person, the high energy intensity of the gross regional product is due to the technological backwardness of the region's energy system and the need to increase the power supply of the regional economy. In the case of unit energy consumption over 8 tce / person, the high energy intensity of the region's gross product is already caused by the low realization of the energy saving potential. According to the latest data from the Federal State Statistics Service of Russia, the energy intensity of

the country's gross product is 14 kg of fuel equivalent per thousand rubles.

Figure 2 shows the distribution of the country's regions in terms of the specific costs of fuel and energy resources and the specific energy intensity of GRP [10, 14, 15].

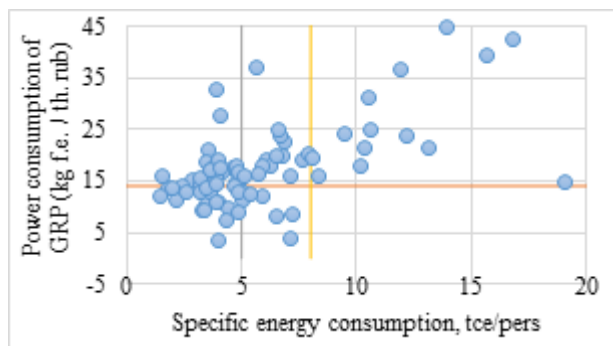


Fig. 2. Distribution of regions of the country by specific costs of fuel and energy resources and specific energy intensity of GRP.

A significant drawback of the above approach to the analysis of the energy intensity of regional power systems is the lack of consideration when grouping the presence of significant differences in the production structure and operating conditions of power systems, which does not allow forming accurate judgments about the reasons for their high energy intensity.

The grouping of territorial energy systems should be based on the following key criteria characterizing the features of the energy supply of the regions:

- structural external (presence or absence of external relations),
- internal structural (structure of available energy resources, production and consumption of energy products),
- balance (energy-deficient, energy-sufficient, energy-surplus in terms of energy resources and products),
- climatic (climatic regions and sub-regions).

To group regional power systems according to a given model of criteria, it is proposed to use data mining methods based on interrelated processes of classification and clustering of initial information.

Methods focused on analyzing the structure of a set of objects form a set of methods for multidimensional classification. Methods of multidimensional classification allow to group objects taking into account essential structural and typological characteristics and the nature of the distribution of objects in a given system of criteria. Classification is carried out on the basis of inclusion of similar objects in a group, and so that objects from different groups are as different as possible [16, 17].

Let all selected m criteria be quantitative. Then each of the n objects can be represented by a point in the m -dimensional criteria space. The nature of the distribution of these points in the space of criteria determines the structure of the similarities and differences of objects. The similarity of objects can be judged by the distance

between the corresponding points. The meaning of this definition of similarity means that objects will be closer when they have fewer differences between the values of the same criteria.

To determine the proximity of a pair of points (objects i and j) in the multidimensional space of criteria, the Euclidean distance is used:

$$d(i, j) = \sqrt{\sum_{k=1}^m (i_k - j_k)^2} \quad (1)$$

Criteria with a large variation range play a large role in calculating the distance between objects. For this reason, the values of the criteria for each object must be standardized:

$$i_k' = \frac{i_k - M[I]}{\sigma_I} \quad (2)$$

where $M[I]$ - expectation operator, σ_I - standard deviation.

Distances for all pairs of objects form a square symmetric distance matrix $D = (d_{ij})^{m,m}$.

The most common methods of multivariate classification are agglomerative-hierarchical and k -means methods.

The agglomerative-hierarchical method is based on the sequential combination of grouped objects. First, the closest objects are combined, and then increasingly distant from each other. The procedure for constructing a classification consists of sequential steps, at each of which two closest groups of objects (clusters) are combined with common properties [18, 19]:

1. A pair of objects is determined, the distance between which is minimal. These objects are combined into one group (cluster), the row and column corresponding to the first of these objects are deleted in the matrix D , and the distances from the new cluster to all other clusters (objects) are calculated as the average of the distances from the objects of the first cluster to all the others. The obtained values are entered into the row and column of the distance matrix corresponding to the second object from the first cluster.

2. Based on the distance matrix reduced by a row and a column, the minimum distance is again determined and a new cluster is formed. This cluster can be built by combining either two objects or one object with the cluster built in the first step.

Usually, the proximity of two clusters is defined as the average value of the distance between all pairs of objects, where one object of a pair belongs to one cluster and the other to another [19, 20].

The agglomerative-hierarchical method provides for the execution of $(n-1)$ iterations, after each of which the number of clusters is decreased by one, and the distance matrix is decreased by a row and a column. At the end of this procedure one cluster that unites all n objects is achieved.

The results of applying the agglomerative-hierarchical method are graphically represented in the

form of a dendrogram (tree of a hierarchical structure) containing n levels, each of which corresponds to one of the steps in the process of sequential cluster enlargement.

Another method of multivariate statistical analysis is the k -means method. In contrast to the agglomerative-hierarchical method, which does not require a preliminary estimate of the possible number of groups of objects, this method is based on the hypothesis of the most probable number of classes. The goal of the method is to build a given number of clusters that should differ as much as possible from each other. The classification building procedure consists of the following steps [20, 21]:

1. A random grouping of objects is carried out and clusters are built.

2. An iterative process of moving objects between groups is carried out in order to minimize the intraclass variance of indicators and maximize the interclass variance (in other words, each cluster should consist of the most similar objects, and the clusters themselves should be different from each other).

The results of this method make it possible to obtain the centers of all classes for each of the initial criteria, as well as to obtain a graphical representation by which criteria the obtained classes differ.

It is proposed to use the agglomerative-hierarchical method to estimate the expected number of groups of regional power systems, and to use the k -means method to assign them to the corresponding groups.

4 Results and discussion

Tables 2 and 3 show the results of grouping the power systems of the country's regions. In total, 8 groups of territorial power systems have been identified, according to which the power systems of 85 regions have been grouped.

Table 2. Results of clustering regional power systems.

Power system criteria system			Region groups									
			1	2	3	4	5	6	7	8		
External relations	absent		+	+								
	available				+	+	+	+	+	+	+	+
Fuel type and origin	transported fuels	gas			+	+	+					
		coal										
	local fuels	gas	+	+						+		
		coal	+	+						+	+	+
	nuclear fuel							+				
	hydropower		+		+			+	+			+
Production scheme	combined		+	+	+	+	+	+	+	+	+	+
	separate			+		+			+	+		
Production capacity	excess		+	+		+	+			+		
	lack				+					+		+
Heat demand	moderate				+	+	+					
	high		+	+						+	+	+
Power consumption density	low				+	+						+
	high		+	+				+	+	+		

Table 3. Results of classification of regional power systems.

Group	Type	Subjects	Power system group characteristics
1	isolated	3	Energy systems of regions with extensive use of hydropower
2		2	Energy systems of regions with pronounced large power generation on fossil fuels
3	open	19	Energy systems of the regions of the non-production sphere of the European part of the country
4		23	Energy systems of post-industrial regions of the European part of the country
5		15	Energy systems of industrial regions of the European part of the country
6		11	Energy systems of energy-intensive industrial regions of the Asian part of the country
7		7	Energy systems of industrial regions of the Asian part of the country with a declining production scale
8		5	Energy systems of the regions of the non-production sphere of the Asian part of the country

The grouping made it possible to compile generalized characteristics of the power systems of isolated and open regions of the country.

So, groups 1 and 2 represent the types of power systems of territorially isolated regions (5 subjects). The common features of power systems in isolated regions are high power density and heat demand. They are characterized by excess production capacity, widespread use of district heating and the use of local energy resources. The dividing line is the use of renewable energy sources (primarily hydropower) in electricity supply. With its development, the share of thermal power plants producing only electricity is significantly reduced.

The power systems of open regions (80 subjects) are characterized by a much broader typification, represented by groups 3 - 8:

- energy systems with fossil fuel energy sources and (or) small cascades of hydroelectric power plants, scarce and poorly diversified in terms of energy resources and insufficient energy capacities with moderate heat consumption and low density of the power supply schedule (regions of the non-production sphere of the European part of the country) (19 entities),
- energy systems with a fossil fuel energy source, sufficient and poorly diversified in terms of energy resources and excess energy capacities with moderate heat consumption and low density of the power supply schedule (post-industrial regions of the European part of the country) (23 entities),
- energy systems based on nuclear and (or) hydropower with the predominant use of fossil fuels in heat supply systems, diversified by energy resources and excess energy capacities with moderate heat consumption and high density of the power supply schedule (industrial regions of the European part of the country) (15 entities),

- energy systems with large power generation on fossil fuel (and hydropower), surplus and poorly diversified by energy resources and insufficient energy capacities with high heat consumption and high density of the power supply schedule (energy-intensive industrial and raw material regions of the Asian part of the country) (11 entities),

- energy systems with fossil-fueled power generation, sufficient in terms of energy resources and excess energy capacities with high heat consumption and high density of the electricity supply schedule (industrial regions of the Asian part of the country with a declining scale of production) (7 entities),

- energy systems with fossil fuel power generation and (or) small cascades of hydroelectric power plants, scarce and poorly diversified in energy resources and insufficient energy capacities with high heat consumption and low density of the power supply schedule (regions of the non-production sphere of the Asian part of the country) (5 entities).

Among the common features of regional power systems should be made:

- strong dependence on fossil fuels (present in the energy balances of all regions and in more than 88% of them it is the main one),

- weak diversification of energy resources (due to the uneven distribution of them, 90% of which are in the Asian part of the country; at the same time, about 86% of natural gas production is exported outside the Asian part and is used in the energy supply of European regions (62%); hence, the main, and often actually the only the energy resource for European regions is natural gas, while most of the Asian regions use coal),

- a heterogeneous industrial structure with a pronounced combined production of energy products (the main production of energy products is concentrated in thermal power plants (66%), but at the same time in the European part they are complemented by nuclear and, partially, hydropower facilities, in Asian regions - large hydroelectric power plants; combined production is one of the key in 97% of the regional power systems),

- shortage (70%) or redundancy in production capacity (due to the presence of external relations in most of the power systems of the regions (92%) and their inertia, which does not allow them to quickly adapt to changing consumption patterns),

- moderate or high demand for heat energy and a gradual decompaction of the electric load schedule (while a significant demand for heat is due to cold climatic conditions, the decompaction of the electric load is caused by a decrease in the share of industry (and partly its energy intensity) in the structure of energy consumption in regions; at the moment, industry remains main consumer in 67% of regions).

Table 4 shows the correlation of the strategies for reducing the energy intensity of the GRP with the groups of regional power systems identified as a result of the clustering.

Table 4. Correlation of strategies for reducing the energy intensity of the GRP with the groups of regional power systems identified as a result of cluster analysis.

Gro up	Priority strategy for reducing the energy intensity of GRP
1	Mastering energy efficient production technologies in order to improve the power system efficiency
2, 5, 6	Reducing energy losses and ensuring resource conservation in the energy sector
3, 8	Mastering energy efficient production technologies that increase the efficiency of the power system in conditions of uneven energy consumption.
4	Reducing the energy intensity of production due to its restructuring and development of the service sector
7	Decommissioning of old and inefficient production equipment and optimization of power system operating modes

As follows from the analysis, for regions with isolated power systems, the choice of the priority strategy for reducing the energy intensity of GRP is largely determined by the structure of the energy carriers used, primarily, by the use of renewable energy sources. In turn, for regions with open energy systems, the choice of a strategy to reduce energy intensity depends on their economic base for sustainable growth and the state of the power infrastructure.

5 Conclusion

Due to the wide variety of energy systems in the regions, their classification is required for their study in order to draw up effective strategies to reduce energy intensity. At the same time, the existing methods of grouping do not take into account the presence of significant differences in the production structure and conditions of operation of power systems, which does not allow forming accurate judgments about the reasons for their high energy intensity and, accordingly, allow a significant classification error.

These differences made it possible to take into account the developed method of grouping the power systems of the regions, which is based on a multivariate statistical analysis of their structural properties and operating conditions using data mining tools.

The developed method for isolated and open regional power systems made it possible to highlight the general features of their functioning and development.

The common features of power systems in isolated regions are high power density and heat demand. They are characterized by excess production capacity, widespread use of district heating and the use of local energy resources.

Among the general features of the power systems of open regions, the following are highlighted: a strong dependence on fossil fuels, weak diversification in energy resources, heterogeneity of the production structure with a pronounced combined energy production, shortage (70%) or redundancy of production capacities, moderate or high demand for thermal energy, gradual decompaction of the schedule electrical load.

For each selected group of regional power systems, priority strategies for reducing energy intensity have been proposed, which will allow achieving the greatest resource saving effect.

This work was performed under the grant of RF President for governmental support of young scientists PhD and DSc (MD-914.2020.6).

References

1. E. Lisin, G. Kurdiukova, N. Ketoeva, J. Katina, Sustainability issues of territorial power systems in market conditions, *Entrepreneurship and sustainability issues*, **6** (2), 1041-1052 (2018)
2. M.E. Bildirici, T. Bakirtas, The relationship among oil, natural gas and coal consumption and economic growth in BRICTS (Brazil, Russian, India, China, Turkey and South Africa) countries, *Energy*, **65**, 134-144 (2014)
3. P. K.Narayan, S. Popp, The energy consumption-real GDP nexus revisited: Empirical evidence from 93 countries, *Economic Modelling*, **29** (2), 303-308 (2012)
4. E. Lisin, W. Strielkowski, V. Chernova, A. Fomina, Assessment of the territorial energy security in the context of energy systems integration. *Energies*, **11** (12), 3284 (2018)
5. A.N. Alekseev, A.V. Bogoviz, L.P. Goncharenko, S.A. Sybachin, A critical review of Russia's energy strategy in the period until 2035, *International Journal of Energy Economics and Policy*, **9** (6), 95-102 (2019)
6. S.V. Ratner, P. Ratner, Regional energy efficiency programs in Russia: The factors of success, *Region*, **3** (1), 71-87 (2016)
7. V. Zybatov, Strategic planning of energy-efficient development of a region of the Russian federation, *Economy of region*, **1** (3), 941-954 (2018)
8. A.A. Makarov, T.A. Mitrova, Strategic development outlook for the energy complex of Russia, *Studies on Russian Economic Development*, **29** (5), 514-526 (2018)
9. G.R. Khasaev, V.A. Tsybatov, *Energy-Efficient and Energy-Safe Development of the Constituent Entities of the Russian Federation: Restrictions and Prospects*, In *Institute of Scientific Communications Conference* (Springer, Cham, 714-727, 2019)
10. *Russian Statistical Yearbook 2018* (Rosstat, Moscow, Russia, 694, 2018)
11. D. Azhgaliyeva, Y. Liu, B. Liddle, An empirical analysis of energy intensity and the role of policy instruments, *Energy Policy*, **145**, 111773 (2020)
12. F.L. Liu, B.W. Ang, Eight methods for decomposing the aggregate energy-intensity of industry, *Applied Energy*, **76** (1-3), 15-23 (2003)
13. S. Voigt, E. De Cian, M. Schymura, E. Verdolini, Energy intensity developments in 40 major economies: structural change or technology improvement?, *Energy Economics*, **41**, 47-62 (2014)
14. L.Y. Bogachkova, S.G. Khurshudyan, Quantitative analysis of energy efficiency indices in the regions of the Russian Federation as exemplified by energy consumption, *International Journal of Energy Economics and Policy*, **5** (4), 1033-1041 (2015)
15. *Regions of Russia, Socio-economic indicators 2017, Statistical collection* (Rosstat, Moscow, Russia, 1404, 2017)
16. V.N. Volkova, V.N. Kozlov, V.E. Mager, L.V. Chernenkaya, Classification of methods and models in system analysis, In *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*, IEEE, 183-186 (2017)
17. G.A. Susto, A. Cenedese, M. Terzi, Time-series classification methods: Review and applications to power systems data, *Big data application in power systems*, 179-220 (2018)
18. M. Roux, Comparative study of divisive and agglomerative hierarchical clustering algorithms, *Journal of Classification*, **35** (2), 345-366 (2018)
19. A. Triayudi, I. Fitri, Comparison of parameter-free agglomerative hierarchical clustering methods, *ICIC Express Letters*, **12** (10), 973-980 (2018)
20. M. Reddy, V. Makara, R.U. Satish, Divisive Hierarchical Clustering with K-means and Agglomerative Hierarchical Clustering, *International Journal of Computer Science and Technology (IJCSST)*, **5** (5), 5-11 (2017)
21. A. Borodkin, E. Lisin, W. Strielkowski, Data Algorithms for Processing and Analysis of Unstructured Text Documents, *Applied Mathematical Sciences*, **25**, 1213-1222 (2014)