

IDENTIFICATION AND CLASSIFICATION OF THE STATES OF COGENERATION SYSTEMS BY COMPETITIVENESS LEVELS OF POWER GENERATING COMPANIES

ALEXEY Y. DOMNIKOV, MICHAEL KHODOROVSKY & LIUDMILA DOMNIKOVA

Academic Department of Banking and Investment Management,
Ural Federal University named after the First President of Russia B.N. Yeltsin, Russia

ABSTRACT

This paper tackles the issues of elaborating and researching methods for generating and selecting information for classifying the states of energy cogeneration systems according to levels of competitiveness. Assessment of the level of competitiveness consists in analyzing the state of centralized and distributed energy cogeneration systems, both in terms of local and integrated indicators of their operations. Joint analysis of the said systems is a complex task that requires the use of adapted methods for its solution. Creating methods for classifying the states of energy cogeneration systems according to levels of their competitiveness is a multilayered task. Since these systems are a combination of engineering, economic, environmental and other subsystems, the number of factors affecting their operations is rather high. A change in the state of energy cogeneration systems can be accompanied by a change in a variety of performance indicators. The behavior of such systems is effectively studied by means of mathematical statistics methods. It is well-known that the issues of diagnosing and modeling the expected level of competitiveness of power generating companies are poorly formalized multiparametric problems with insufficiently defined information and multidimensional relationships between indicators that typify competitiveness. Therefore, the procedure for classifying the states of energy cogeneration systems according to their competitiveness levels, with subsequent review of both the properties of individual classes and the differences between them, fit well into the discriminant analysis model. The use of discriminant analysis technique made it possible to build a classification system that allows to identify the state of multidimensional objects in energy cogeneration systems.

Keywords: power generation industry, energy cogeneration systems, classification, discriminant analysis, competitiveness, uncertainty.

1 INTRODUCTION

The reliable operation of energy cogeneration systems, being the most important component of major power generation systems, is essential for successful economic development of the country as a whole. Not only technical, but also economic aspects of reliability predetermine the complexity of studying the above-mentioned objects and their interaction with other components of economy and social sphere in order to determine the best control actions to achieve an economic effect and to maintain a constant readiness of power systems to cope with operational threats arising in periods of economic and political crises, catastrophes, disasters, etc. The complex structure of energy cogeneration systems requires the development of adequate methods for determining their states.

The tasks to be solved in the study comprise the creation and review of methods for generating and selecting information for classifying the states of energy cogeneration systems according to the levels of competitiveness of power generating companies, as well as the analysis of their capabilities to ensure the reliable power supply.

Assessment of the level of competitiveness of energy cogeneration systems consists in analyzing the state of territorial energy systems, fuel supply systems, and centralized and



local heat supply systems, in terms of their integrated indicators of their operation. Joint analysis of the said systems is a complex task that requires the use of well-proven and adapted methods for its solution. Therefore, this research is not focused on the entire set of energy industry's systems, but only on a fraction of it – cogeneration systems. A cogeneration system is based on the following: (1) centralized system, consisting mainly of heating-type cogeneration sources that form part of the territorial generating company; (2) distributed system, consisting of medium- and low-power cogeneration sources located near power load centers, including those serving individual groups and units of consumers (usually in remote areas of a region). If combined in a rational way, the centralized and distributed systems can complement each other and make it possible to create a sufficiently flexible energy cogeneration system for a territory, with it becoming a key link in the regional power generation system. Such a cogeneration system is able to reliably provide consumers with electric and thermal power and successfully compete with separately generated energy sources due to the implementation of cogeneration's competitive advantages [1], [2].

Thus, creating methods for classifying the states of energy cogeneration systems is a multilayered task. Since the energy cogeneration system (object to be classified) is a combination of engineering, economic and other subsystems, the number of factors affecting its operation is quite large. Their changes are often random. A change in the state of energy cogeneration systems can be accompanied by a change in a variety of performance indicators. Research experience has shown that the behavior of such systems is effectively studied by means of mathematical statistics methods. At the same time, it has to be borne in mind that applying the mathematical statistics methods only formally may produce results that in fact do not completely line up with reality. These results cannot be properly interpreted, and this will lead to confusion and reduce the confidence in the research method used [3], [4].

2 METHODOLOGICAL ASPECTS OF CLASSIFICATION OF STATES OF ENERGY COGENERATION SYSTEMS BASED ON DISCRIMINANT ANALYSIS

In order to classify the states of energy cogeneration systems, we have considered and can recommend the discriminant analysis, as it enables us to build a classification system based on the data of a training sample. The large number of procedures and checks employed in the method allows making a priori conclusions on the resulting classification system's quality. The essence of the method consists in maximizing differences between classes while minimizing intragroup differences based on the training sample's data. Thus, the data and variables used to obtain discriminant functions are to be as representative of the classes being generated as possible [5], [6].

The discriminant analysis method's advantages for analyzing the state of energy cogeneration systems in different territories, compared to other methods, are as follows:

- Analysis model is built on the basis of maximizing the differences between identified classes;
- The model provides for subordination of observations made in classes to a certain probabilistic law (the normality requirement may not be met in case of using special procedures);
- Along with factor analysis, it replaces the space of parameters (variables) with a factor space (which, in the case of the issue being analyzed, has much less dimensions).

The methodological approach that has been developed is as follows. Prior to elaborating a classification system, the initial data analysis is effected: subordination of variables to the normal probability law, equality of covariance matrices of classes, etc.



In general, the task consists in constructing, based on the training sample data (with an estimate of the general totality parameters), the image models, i.e. decision rules that would serve as the basis for recognizing a new object $a(x_0)$:

$$X_0 \in m_k \Rightarrow X_0 \in X_k \Rightarrow a(X_0) \in A(X_k), k = N, M, L. \quad (1)$$

The decision rule has to enable minimizing the mathematical expectancy of losses associated with misclassification, that is:

$$F(u) = c_1 q_1 P_1(u) + c_2 q_2 P_2(u) + c_3 q_3 P_3(u), \quad (2)$$

where $P_k(u)$, ($k = 1, 2, 3$) = functions for distribution of objects in classes (N, M, L); q_k = a priori probabilities of the emergence of objects of three classes; c_k = “costs” of errors in assigning objects to three classes.

It is possible to effect the classification of observations on the basis of measuring the distances between the object and the centroids of classes in the space of canonical discriminant functions [7]. The classical version of discriminant analysis, which determines the canonical directions in the attributes space, is based on the following criterion:

$$D = \frac{\text{interclass variance}}{\text{intraclass variance}} = \max. \quad (3)$$

3 GENERATION OF A CLASSIFICATION SYSTEM AND ASSESSMENT OF ITS QUALITY

After carrying out the preliminary procedure for selecting the most informative indicators and statistically analyzing the data, we turn to the principal part of the analysis – obtaining discriminant functions and identifying new observations presented for identification [8], [9]. The new facilities correspond to the cogeneration systems of the Urals, typified by their year 2020 indicators. The analysis is meant to determine which class the objects belong to, which level of crisis they experience, and how their state has changed in comparison with previous years.

The elaboration of a classification system involves obtaining the coefficients of canonical discriminant functions that determine the system of orthogonal directions that delineate the maximum differences between the classes, be it in a standardized or non-standardized form. Table 1 lists these coefficients.

In order to classify new observations when using functions with coefficients in a non-standardized way, the values of variables are used in their original form. There will be some peculiarities in the values of discriminant functions obtained in this way. In particular, the variables have unequal weight, and their contribution to values of discriminant functions is calculated in absolute (actual) units. This leads to the fact that coefficient values in case of some variables will be rather large, while in case of others – rather small. Indeed, the values of coefficients (in absolute magnitude) in Table 1 for variables 10 and 12 are four orders of magnitude higher than the value of coefficient for variable 14 in both discriminant functions. However, this does not correspond to the actual contribution of variables to values of discriminant functions.

When using standardized coefficients, the contribution of each variable is measured in relative units, i.e. in the form of the variable’s contribution to the discriminant function’s value. The standardized variables have a zero mathematical expected value and an occasional root-mean-square deviation [10].



Table 1: Coefficients of canonical discriminant functions.

Variable name	Canonical discriminant functions			
	Non-standardized		Standardized	
	Function number		Function number	
	1	2	1	2
1. Territory's operating capability to peak load ratio	−0.979	3.615	−0.278	1.497
2. Installed capacity use factor	−1.671	10.874	−0.221	1.441
3. Share of cycling power sources in the installed capacity	4.833	−2.482	0.393	−0.202
4. Ratio net power flow to electricity generated in the territory	−0.240	−0.329	−0.374	−0.512
5. Power consumption per person	0.146	−0.651	0.271	−1.209
6. Installed capacity use factor of captive power plants	2.343	−9.162	0.565	−2.211
7. Share of electricity generated by captive power plants in total power generation	−2.329	16.368	−0.256	1.797
8. Normalized heat generation per one kilometer of main pipelines	0.020	0.007	0.653	0.220
9. Share of fuel oil in the total reference fuel consumption	−3.426	−7.344	−0.255	−0.546
10. Share of natural gas in the total reference fuel consumption	1.913	1.671	0.401	0.350
11. Share of other types of fuel in the total reference fuel consumption	−2.306	14.407	−0.090	0.564
12. Specific losses in networks, per unit of length of transmission line normalized to 110 kV	38.346	−9.126	1.105	−0.263
13. Ratio of length of power lines normalized to 110 kV to the total power consumption in the territory	0.570	−2.041	0.370	−1.323
14. Normalized fuel consumption for power generation	−0.002	0.009	−0.131	0.569
15. Normalized fuel consumption for heat generation	0.041	0.012	0.633	0.181
16. Ratio of power grid losses to the total of electricity consumed without accounting for the losses	−28.745	38.887	−0.745	1.008
17. Number of process or functional failures per kW of installed capacity	0.002	0.006	0.162	0.557
18. Number of process or functional failures per thousand kilometers of power transmission lines normalized to 110 kV	−0.058	−0.148	−0.403	−1.026
Constant	−10.614	−5.856	0	0

Comparing the coefficients of discriminant functions specified in Table 1, note the concordance of signs for the corresponding non-standardized and standardized coefficients, which should be the case. Judging by the values of standardized coefficients, we note that the variable 12 makes the largest contribution to the first function's value. Variable 11 makes the smallest contribution, moreover, this contribution is negative. The largest negative contribution is made by variable 16, and the largest positive contribution to the second function's value is made by variable 7, while the largest negative contribution is made by 6, which is greater in absolute value. Variable 8 makes the smallest contribution.

The constants in the discriminant functions are calculated in such a way that the principal centroid has zero coordinates in the discriminant functions plane. This condition is satisfied automatically for any discriminant functions in a standardized form, so the constants are equal to zero.

The centroids of classes are calculated by means of replacing the variables with their average values according to groups. This results in three points corresponding to the centers of classes, or the so-called centers of "gravity" of the classes (Table 2).

Table 2: Values of functions at centroid points of classes.

Competitiveness level	Function	
	1	2
<i>N</i>	3.793	0.835
<i>M</i>	0.456	−1.306
<i>L</i>	−1.299	0.455

The identification of objects is effected as follows. The object to be identified has its distances to the centers of all classes calculated. The object is assigned to a class to which the distance is the shortest. Thus, a classification system was obtained based on discriminant analysis of a training sample data on 90 observations corresponding to 18 variables.

In addition to identifying the objects, their position in the plane of discriminant functions offers additional information for analysis: proximity of object to class boundaries, dynamics of object situation change and trajectory of their "movement", since there is a constant set of objects (eight energy cogeneration systems) that is tracked over time.

Before proceeding to the analysis of results, let us consider the assessment of the resulting system's quality. Tables 3 and 4 list the criteria's computed values.

Table 3: Values of utility criteria for canonical discriminant functions.

Function	Eigenvalue λ	Relative percentage, %	Canonical correlation
1	3.347	82	0.887
2	0.736	18	0.651

Table 4: Assessment of residual discriminant capability.

Functions	λ -Wilks statistics	χ^2 test of significance	Number of degrees of freedom	Significance
1	0.133	158.639	36	0.0001
1 and 2	0.576	43.28	17	0.0001



The eigenvalues of discriminant functions and their relative percentage in description of the total variance show how the discriminant functions relate to each other in terms of shared properties. The relative percentage for the first function is noticeably higher than for the second. However, there is not much difference in the canonical correlation values.

λ -Wilks statistics and significance test also demonstrate the presence of significant residual discriminant capability. A significance of 0.0001 suggests that if there are no differences between classes in the unused part of the data variation, then such a large statistic value is found in one out of 10,000 samples. Therefore, this classification system can be used for identification of objects and their subsequent analysis.

4 ANALYSIS OF RESULTS OF CLASSIFICATION OF THE URALS' ENERGY COGENERATION SYSTEMS

The input of the classification system receives descriptions of classified objects according to values of integral indicators as of the end of 2020 – Bashkir, Kurgan, Orenburg, Perm, Sverdlovsk, Udmurt, Chelyabinsk and Tyumen cogeneration systems of the Urals region. Table 5 lists the results of classification of these objects for the year 2020.

Table 5: Object identification results (2020).

Energy cogeneration system	Greatest of class membership probabilities				Greatest of probabilities of belonging to other classes			Discriminant function value		Competitiveness level
	Identified	Prior probability	Posterior probability	Generalized distance	Group	Posterior probability	Generalized distance	1	2	
Bashkir	2	0.11285	0.83767	4.36346	3	0.16233	7.64546	-1.37099	-2.31911	M
Kurgan	3	0.07346	0.79908	5.22211	2	0.20092	7.98328	-2.35647	-1.58081	L
Orenburg	3	0.43645	0.57953	1.65818	2	0.42046	2.29991	-0.97462	-0.80117	L
Perm	3	0.89291	0.95768	0.22653	2	0.04231	6.46571	-0.97681	0.79532	L
Sverdlovsk	2	0.00001	0.98150	23.61204	3	0.01850	31.55461	-2.71000	-4.99225	M
Udmurt	3	0.09616	0.77603	4.68350	2	0.22397	7.16883	-2.21359	-1.51639	L
Chelyabinsk	3	0.44567	0.99799	1.61636	2	0.00201	14.03158	-2.10670	1.42683	L
Tyumen	2	0.00001	0.99999	23.31339	3	0.00001	46.59361	0.52200	-6.13357	M

As a result of classification, the calculated level of competitiveness of cogeneration systems receives the following designation: N = Normal, M = Medium, L = Low.

The analysis of results showed that the Bashkir, Sverdlovsk and Tyumen cogeneration systems are classified as corresponding to M class, and the rest of the classified objects are classified as corresponding to L class. If we consider the values of a posterior probabilities for three objects from the class M – the Bashkir, Sverdlovsk and Tyumen energy cogeneration systems – then they are 0.83767, 0.981499 and 0.999991, i.e. close to one. The

values of largest a posterior probabilities of these observations belonging to other classes (class L) are, respectively, 0.162329, 0.0118501, and 8.81×10^{-6} . As for the rest of energy cogeneration systems, their class is also unambiguous (value of the largest posterior probability is much higher than the other posterior probabilities; values of posterior probabilities of objects of training sample also differ greatly in case of the nearest two classes). Only in the case of Orenburg energy cogeneration system, the posterior probability of belonging to the M and L classes amounts to 0.420462 and 0.579534, respectively. Obviously, in terms of the plane of discriminant functions this object is at approximately the same distance from classes M and L, i.e. it is close to the class boundary. The Bashkir, Kurgan and Udmurt cogeneration systems are also located near the M and L classes' boundary. The Sverdlovsk, Chelyabinsk and Tyumen cogeneration systems are located at a greater distance from the M and L classes' boundary.

Based on the calculations performed, some general remarks can be made on the applicability of discriminant analysis for the issue being considered. First, there is a fairly good agreement of results with initial data (training sample) on class affiliation. Second, the objects' position on the discriminant functions plane confirmed that, for example, the Perm cogeneration system is in a worse position in terms of competitiveness than the Sverdlovsk one, i.e. the position of objects in the model reflects actual objective properties.

5 CONCLUSION

Based on the 2010–2020 data on energy cogeneration systems in the Urals, a classification system was obtained, making it possible to determine the state of electric power systems and diagnose the situation for each facility by examining the variables included in the calculation. Information on the situational dynamics of various objects has also been obtained, which may be useful for short-term forecasting of states of reviewed energy cogeneration systems.

The advantage of the proposed methodological approach to classification and identification of states of energy cogeneration systems consists in enhancement of decision-making technology pertaining to minimization of economic risks in a situation of uncertainty, aimed at increasing a power generating company's competitiveness.

ACKNOWLEDGEMENTS

The work was supported by Act 211 Government of the Russian Federation, contract No 02. A03.21.0006 and Russian Foundation for Basic Research (RFBR), contract № 20-010-00886.

REFERENCES

- [1] Domnikov, A., Antipova, E. & Domnikova, L., Diagnostics of competitiveness of power-generating companies. *WIT Transactions on Ecology and the Environment*, vol. 222, WIT Press: Southampton and Boston, pp. 27–33, 2019. DOI: 10.2495/EQ180031.
- [2] Domnikov, A., Antipova, E. & Domnikova, L., Evaluation of economic risks for power-generating companies. *WIT Transactions on Engineering Sciences*, vol. 121, WIT Press: Southampton and Boston, pp. 125–133 2019. DOI: 10.2495/RISK180111.
- [3] Bolch, B. & Huang, C., *Multivariate Statistical Methods for Business and Economics*, Prentice-Hall: New Jersey, 1973.
- [4] Mehigan, L., A review of the role of distributed generation (DG) in future electricity systems. *Energy*, **163**, pp. 822–836, 2018. DOI: 10.1016/j.energy.2018.08.022.
- [5] Bai, J. & Perron, P., Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, **18**, pp. 1–22, 2003.



- [6] Lance, G. & Williams, W.T., A general theory of classificatory sorting strategies: 1. Hierarchical systems. *The Computer Journal*, **9**(4), pp. 373–380, 1967.
- [7] Lachenbruch, P., *Discriminant Analysis*, Hafner Press: New York, 1995.
- [8] Hayek, F., The meaning of competition. *Austrian Economics*, Hillsdale: Michigan, pp. 264–280, 1991.
- [9] Stevens, J., *Applied Multivariate Statistics for the Social Sciences*, Lawrence Erlbaum: New Jersey, 1996.
- [10] Greene, W., *Econometric Analysis*, Macmillan: New York, 2003.

