ORIGINAL SCIENTIFIC RESEARCH

THE EXPERIENCE OF APPLYING MATHEMATICAL METHODS FOR ANALYSIS OF THE MICROGENERATION SECTOR IN RUSSIA

VLADIMIROVICH Gavriusev Sergei [1], VASILYEVICH Chernyaev Maxim [2]

[1]*Peoples' Friendship University of Russia (RUDN University), Moscow (RUSSIA)*
[2]*Peoples' Friendship University of Russia (RUDN University), Moscow (RUSSIA)*
E-mail:1042180140@rudn.ru , chernyaev-mv@rudn.ru

ABSTRACT

The paper relates to substantiation of reasonability of applying the mathematical methods to economic researches. The developed by the author algorithm for determining the relevant for Russia foreign experience of the microgeneration sector development on the basis of the Gaussian mixture model is provided as an example. This algorithm used to be successfully applied within the framework of a larger scale research related to the problems of development and increasing of the economic efficiency of the microgeneration sector in Russia. The paper provides for the mathematical description and economic feasibility study of the measures taken. High efficiency of using the quantitative methods for obtaining the results, which are significant from the point of view of the economic feasibility study, is underlined. There are also described the problematic issues of the suggested algorithm and possible ways of solving thereof. Finally, the conclusion is made about a high heuristic value of the above-described algorithm and the reasonability of applying the mathematical methods to economic researches.

Keywords: microgeneration sector, Gaussian mixture model (GMM), clustering, mathematical methods, electricity sector.

IMPORTANCE OF INVESTIGATION OF THE MICROGENERATION SECTOR IN RUSSIA

The electric power generation and distribution system of the Russian Federation is currently operating based on the principle of the united energy system with domination of large-scale suppliers of the electric power. Despite the fact that electric grids have long ranges, they cover 30 - 40 % [13] of the territory, wherein about 20 % [13] of the population have no possibilities to get connected to the centralized grid electricity. The global trend to decarbonization of the world economy is an additional and rather powerful issue exerting the influence upon the electric power generation and distribution system of Russia. The trend is still more definitely shaped if regarded from the point of view of the market regulation. The most recent regulatory and compliance initiatives of the European Union (EU), a special place within which is occupied by introduction of the carbon border tax, confirm distribution of the environmentally determined regulation to the interstate (transboundary) level. Development of the microgeneration sector might, as it seems, turn out to be a pro factor for adaptation of the Russian electric power generation and distribution system to the new regulatory environment.

SUBSTANTIATION OF APPLICATION OF THE MATHEMATICAL METHODS

The necessity in availability of the results subject to verification obtained on the basis of the methods to be proved determines strive for using quantitative methods in the research. As an example, we can provide the algorithm for determining relevant for Russia foreign experience of the microgeneration sector development based on the method using the Gaussian mixture model (GMM). Within the framework of the research related to development and increasing of the economic efficiency of the microgeneration sector in Russia a mission was set to investigate foreign experience. For the purpose thereof, it was necessary to find the countries, the experience of which would be the most useful for the conditions of Russia. Mathematical problem of selection of the proper countries was reduced to the clustering problem that was solved with the help of applying the GMM. This paper is related to description of application of the present method for solving the practical problem within the framework of economic research as well as to analysis of the results of using the above mathematical method.

The objective of the present paper is in showing the reasonability of applying the GMM as an economic research tool based on the example of the algorithm for determining the relevant for Russia foreign experience of the microgeneration sector development.

METHODS

Using of the foreign experience for analysis of the microgeneration sector in Russia is, on the one hand, an important component of the research, and, on the other hand, it requires that the examples would be selected thoroughly. Images of the countries under consideration were formed in order to select the countries, the electric power generation and distribution systems of which could be considered as the relevant analogs. The idea of forming the images was derived from the fact that within the framework of the performed studying of the microgeneration sector it was important to consider not only the electric sector component, but also the conditions, which were exogenous for this system, like economic and social and demographic issues, within which the electricity sector was developed. Due to the fact that within the framework of the performed study of the microgeneration sector the focus of scientific interest was directed to the perspectives of development of this kind of the electricity sector in Russia, the decision was taken to consider the image of Russia "in the future". This determined formation of the image of Russia not based on current values of the basic parameters, but on the basis of the base scenario of the forecast of the social and economic development and the development of the energy system of the country. The year of 2035 was selected as the time limit because it was the year, for which all the necessary data were available. Images of other countries were formed on the basis of the open source data taken during the period from 1991 up to 2019. That happened due to the fact that we managed to find out all the required for the analysis data for all those years and also because the Russian Federation was established as an independent state in 1991.

There were considered the following parameters forming the image of the country:

1. amount of the electric power generated by the electricity sector of the country annually;
2. gross domestic product (GDP) of the country;
3. number of population;
4. median of the distances between neighboring cities;
5. share of the electric power generated by hydro power plants;
6. share of the electric power generated by atomic power plants;
7. share of the electric power generated on the basis of the combustible fuel;
8. share of the electric power generated on the basis of the renewable energy sources (RES).

The following were used as the data sources for formation of the image of Russia. The annually generated power parameters of the electric power generation and distribution system of the country, as well as the shares thereof taken upon the types of fuel were obtained from the "Energy strategy of Russia for the period up to 2035" [16]. The GDP parameters were taken from the "Forecast of social and economic development of the Russian Federation for the period of up to 2036" [15]. The data related to the projected population were taken from the bulletin "Projected number of population of the Russian Federation by 2035" [14].

Images of other countries were formed on the basis of the information obtained from the open databases. Annual generation of electric power per country, aggregate shares of electric power plants operating on the basis of various kinds of fuel, as well as the number of population in the country were taken from British Petroleum Stats Review 2020 Consolidated dataset Panel format [3]. Data about the GDP of other countries were taken from the statistical data prepared by the World Bank "GDP (current US$)" [5].

Medians of the distances between the nearest cities for both Russia and for other countries were obtained with the help of processing of the information taken from the database source called "World Cities Database" [12].

The Gaussian mixture model was selected as the clustering algorithm. The selection is stipulated by the fact that it is possible to calculate the optimal number of the clusters, to get access to the settings of the initial parameters and application of the software tools for calculations is also possible. It is important to note that all the calculations performed in the process of this investigation were done with the help of the Jupyter Notebook [1] and PyCharm Community [4] software using the distributive of the Python Anaconda programming language [2], along with the NumPy [6], Scikit-learn [8], Matplotlib [7], and Plotly [9] libraries. Despite the fact that in the general case the GMM works quite efficiently with the input data of different orders, its application for solving of the clustering problem requires some specific additional manipulations in order to decrease the calculation error. This is stipulated by the algorithm for selection of the relevant examples. The following steps were taken in order to accomplish the given problem:

1. preparation of the data;
2. selection of the best year for each country as compared with the image of Russia;
3. solving of the clustering problem.

Preparation of the data included, primarily, minimization of the difference in the orders of the values of input parameters. The following was done in this relation. First, the generated electric power was expressed in TW, GDP – in trillion US dollars and the number of population – in million people. Second, all the data except for the shares of the electric power plants operating on the basis of various types of fuel were reduced to the logarithmic notation.

Let us introduce some designations in order to describe step two:

$Countries\{\}$ is the set of all the countries;

country is a separate country: $country \in Countries$;

$Years\{\}$ is the set of all the analyzed years:

$Years = [1991; 2019]$;

year is a separate year: $y \in Years$;

Parameters{}is an ordered set of all the parameters;

p is a separate parameter: $p \in P$;

$p_{(year)}$is the value of the parameter in a certain year.

The images were created for each country for every year starting from 1991 and up to 2019. Every image was compared with the image of Russia, and the sum of squares of the differences in values from

each pair of the correspondent parameters was calculated. We define as the $year_{best}$ the year, during which the above value appeared to be minimal (Formula 1).

Formula 1

$$year_{best} = \min_{1991 \leq y \leq 2019} \sum_{i=1}^{Card(Parameters)} (p_{i(year)} - p_{rus(i)})^2,$$

$year_{best}$ is the required year, during which the minimum difference is attained,

$p_{rus(i)}$ is the parameter for the image of Russia;

Sources: compiled by the author.

We determine the image of the country passed for the subsequent analysis as $Image_{country}\{\}$ (Formula 2).

Formula 2

$$Image_{country} = \{p_{1(y_{best})}, \dots, p_{Card(Parameters)(y_{best})}\}$$

Solving of the clustering problem was performed as follows.

We create a set of images of all the countries selected at the previous step and designate it as AllCountries{} (Formula 3).

Formula 3

$$AllCountries = \{\forall country \in Countries: Image_{country}\}$$

The efficiency of the GMM application depends primarily upon the number of components and selection of the covariance type. The number of the components, which is rational in this case, lies within the limits of the section from 2 to 12. The accessible for performance of the calculations covariance types are represented by the following ones:

"full" means that eigenvalue covariance matrix is calculated for each of the components;

"tied" means that one covariance matrix is calculated for all the components;

"diag" means that every component possesses its own diagonal covariance matrix;

"spherical" means that the value of dispersion is calculated for each of the components [11].

Optimal combination of the number of components and the covariance types – X{} of such parameters is determined with the help of calculation for each of the information criteria depending on the log likelihood function (Formula 4) – the Bayesian information criterion (BIC) (Formula 5) and the Akaike information criterion (AIC) (k is the number of parameters;

n is the number of observations.

Formula 6). Strictly speaking, we could do it by calculating the BIC only as the criterion, which is stricter to the number of components. But to attain higher accuracy and for the purpose of visualization a decision was made to calculate both of the above criteria.

Formula 4

$$L(X|\theta) = \sum_{i=1}^{Card(X)} \ln(f_x(x_i|\theta)),$$
$f_x(x_i|\theta)$ is the joint probability distribution function for all of the combinations.

Formula 5

$$BIC = k * \ln(n) - 2 * L,$$
L is the maximum value of the log likelihood function;
k is the number of parameters;
n is the number of observations.

Formula 6

$$AIC = 2 * k - 2L,$$
L is the maximum value of the log likelihood function;
k is the number of parameters.

Let us represent the obtained results on the graphs (Figure 1). It follows from the results that in order to obtain the best result we have to divide the sample into ten clusters at the "full" covariance type – performance of the calculation for each component of a separate covariance matrix.
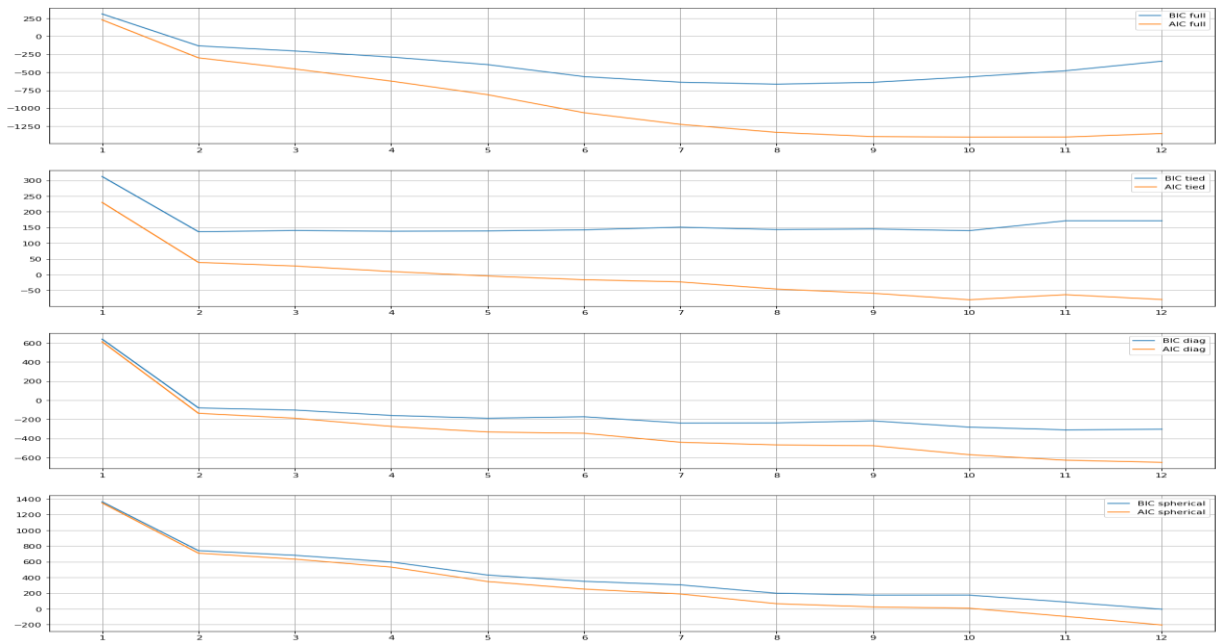
Figure 1

The values of BIC and AIC for the combination of the number of the model parameters and the covariance type.

Sources: compiled by the author based on the open access sources.

* Energy strategy of Russia for the period up to 2035 [16]
* Forecast of social and economic development of the Russian Federation for the period of up to 2036 [15]
* Projected number of population of the Russian Federation by 2035 [14]
* British Petroleum Stats Review 2020 Consolidated dataset Panel format [3]
* GDP (current US$) [5]
* World Cities Database [12]

We apply the GMM.

The GMM algorithm has the following representation (Formula 7).

Formula 7

Let $\ell = \text{Card}(X), n = \text{Card}(P)$

1. $\forall a \in Y$ initial approximations are set $\omega_a, \mu_a, \sum_a$;
2. Repeat until a_i ceases to vary:
   2.1. E-step ($\forall x_i$ are referred to the nearest centers):

$$g_{ia} := P(a|x_i) = \frac{\omega_a p_a(x_i)}{\sum_y \omega_y p_y(x_i)}, a \in Y, i = 1, \dots, \ell;$$

$$a_i := \arg\max g_{ia}, a \in Y, i = 1, \dots, \ell;$$

   2.2. M-step (calculation of new positions of the centers):

$$\omega_a := \frac{1}{\ell} \sum_{i=1}^{\ell} g_{ia}, a \in Y;$$

$$\mu_{ad} := \frac{1}{\ell \omega_a} \sum_{i=1}^{\ell} g_{ia} f_d(x_i), a \in Y, d = 1, \dots, n;$$

$$\sigma^2_{ad} := \frac{1}{\ell \omega_a} \sum_{i=1}^{\ell} g_{ia}(f_d(x_i) - \mu_{ad}), a \in Y, d = 1, \dots, n;$$

Sources: compiled by the author based on the open access sources.
Gaussian mixture models and the EM algorithm [10]

Visualization of the results does not seem possible due to multidimensionality of the clustering. Resulting from the performed calculations, the Russian Federation appeared within the same cluster with the states like the United States of America (the USA) and the Republic of India (India).

Due to the particular features of the electric power generation and distribution market of Russia, as well as considering its geographical position, positioning of industrial facilities and major private consumers, it was also solved the clustering problem for the most economically developed territory of the European part of the Russian Federation. In this case two countries – the United Kingdom of Great Britain and Northern Ireland (Great Britain) and Federal Republic of Germany (Germany) – were included into the same cluster with the designated territory.

RESULTS AND DISCUSSIONS

Basic result of application of the aforementioned algorithm was in the fact that there were selected the countries, the electric power generation and distribution systems of which considering social and economic and demographic particularities had, during a specific period of time, maximum similarity with the expected state of the electricity sector of Russia in the case of implementation of the base scenario of development. India was excluded in the process of further consideration. The reason for such exclusion was in a rapid growth of the amounts of the electric power generation in the country that did not allow use that experience for Russia, development of the electric power generation and distribution system of which possesses a far more steady mode.

Consideration of other countries like the USA, Great Britain and Germany allowed detecting basic trends of development for their electric power generation and distribution systems. There were also determined the measures taken by the governments of the countries concerned for development of the microgeneration sector. The obtained information formed the basis for development of the specific for Russia recommendations related to development and increasing of economic efficiency of the microgeneration sector.

CONCLUSIONS AND RECOMMENDATIONS

The efficiency of application of the GMM comprised in this case 75 % (3 of the 4 selected countries used to be confirmed in the process of further investigation) evidencing for the fact that application of the above mechanism is justified. The algorithm described in this paper and having the GMM as an integral part of it possesses high heuristic value. In this case it was aimed at obtaining of the best result for the Russian Federation but, if considered on general terms, the algorithm can be applied to any country and any time period.

High degree of usefulness of the suggested algorithm does not deprive it from certain drawbacks. The following disadvantages should be noted among the most significant ones. First, the suggested algorithm is calculated for application in one country and during a certain specific moment of time. In the case when there is more than one country for which similar algorithms are to be found, a more complicated system should be applied for estimation of the best similarity. A similar situation occurs in the case when it is necessary to perform analysis for a certain period of time. Using of an auxiliary parameter based on application of the weighting coefficients could be suggested as a possible solution. Applying of a simple average is possible in the case when importance for the analysis of all the countries in the sample and all the years, during which the selection of foreign experience has to be performed, is equivalent. At differentiation of either the countries or the time periods upon their importance for the purposes of the investigation, various mixtures of the weighting coefficients should be applied.

Another vulnerability of the suggested algorithm is in the fact that it does not consider the dynamics of development of the electric power generation and distribution system along with the conditions in the external in relation to it systems (economic, social and demographic parameters of the model). India was sampled erroneously for the above reason. Possible solution is seen in using of a more complicated algorithm considering the value of deviation from the best value ($Image_{country}$). This would allow considering the dynamics of development and preventing from joining into one and the same cluster of the countries, the history of development of which has substantial differences. A similar requirement could be

also set to the analyzed country or a group of countries in a more general case. For the situation like that there occurs an additional complication in the form of the necessity of using the matrix calculation already at the stage of preparation of the data for application of the GMM.

Despite the above-mentioned problematic issues of the considered algorithm, it showed its efficiency and can be applied for economic researches both in the form represented in this paper and in the form to be worked out on the basis of the suggested recommendations. The considered particular case proves availability of broad opportunities for the reasoned application of the mathematical methods to economic researches.

ACKNOWLEDGEMENT

REFERENCES

[1] About us [Electronic resource] // jupyter. – URL: https://jupyter.org/about. – open access. – (date of access: 25.05.2021).

[2] Anaconda [Electronic resource] // Anaconda. – URL: https://www.anaconda.com/products/individual. – open access. – (date of access: 25.05.2021).

[3] BP Stats Review 2020 Consolidated dataset Panel format [Electronic resource] // BP. –URL: https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html. – open access. – (date of access: 25.05.2021).

[4] Download PyCharm [Electronic resource] // PyCharm. – URL: https://www.jetbrains.com/pycharm/download/#section=windows. – open access. – (date of access: 25.05.2021).

[5] GDP (current US$) [Electronic resource] // World Bank. – URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD. – open access. – (date of access: 25.05.2021).

[6] Harris C.R. [et al.] Array programming with NumPy / C.R. Harris [et al.] // Nature. – 2020. – No. 585. – P.: 357-362.

[7] Matplotlib [Electronic resource] // Matplotlib. – URL: https://matplotlib.org/stable/index.html. – open access. – (date of access: 25.05.2021).

[8] Pedregosa F. [et al.] Scikit-learn: Machine Learning in Python / F. Pedregosa [et al.] // Journal of Machine Learning Research. – 2011. – No. 12. – P.: 2825-2830.

[9] Plotly Python Open Source Graphing Library [Electronic resource] // Plotly. – URL: https://plotly.com/python/. – open access. – (date of access: 25.05.2021).

[10] Sridharan R. Gaussian mixture models and the EM algorithm [Electronic article] / R. Sridhara // MIT. – URL: https://people.csail.mit.edu/rameshvs/content/gmm-em.pdf. – open access. – (date of access: 25.05.2021).

[11] Variational Gaussian Mixture Models [Electronic resource] // scikit-learn. – URL: https://scikit-learn.org/0.17/modules/dp-derivation.html. – open access. – (date of access: 25.05.2021).

[12] World Cities Database [Electronic resource] // kaggle. – URL: https://www.kaggle.com/max-mind/world-cities-database/discussion/80303. – open access. – (date of access: 25.05.2021).

[13] Dziubenko S. Remotely heated: Own generation and private investments set the trend for the sector [Electronic article] / S. Dziubenko // Rossiiskaia gazeta. – URL: https://rg.ru/2016/12/21/malaia-energetika-stala-trendom-na-rossijskom-rynke.html. – open access. – (date of access: 25.05.2021).

[14] Projected number of population of the Russian Federation by 2035 [Electronic resource] // Federal State Statistics Service. – URL: https://rosstat.gov.ru/compendium/document/13285?. – open access. – (date of access: 25.05.2021).

[15] Forecast of Social and Economic Development of the Russian Federation for the Period Up to 2036 [Electronic resource] // Federal State Statistics Service. – URL: https://www.economy.gov.ru/material/file/a5f3add5deab665b344b47a8786dc902/prognoz2036.pdf. – open access. – (date of access: 25.05.2021).

[16] Energy strategy of Russia for the period up to 2035 / Ministry of Energy of the Russian Federation. – URL: https://minenergo.gov.ru/sites/default/files/documents/11/10/1920/document-66308.pdf. – open access. – (date of access: 25.05.2021).