

# Development of methods for the formation of operation modes of hydropower systems using machine learning

Ayrat Mardikhanov<sup>1,\*</sup>, Vilen Sharifullin<sup>2</sup>, A V Golenishchev-Kutuzov<sup>2</sup> and Sh G Ziganshin<sup>2</sup>

<sup>1</sup>JSC Tatenergo, 420021, Salimzhanova 1, Kazan, Russian Federation

<sup>2</sup>Kazan State Power Engineering University, 420066, Krasnoselskaya str., 51, Kazan, Russian Federation

**Abstract.** The paper describes the method for finding a compromise solution during formation of operation modes of hydropower systems (cascade of hydropower plants). The software solution “Energy system of the HPP cascade” (<http://hydrocascade.com>) was implemented based on the developed methodology. In the existing model, in order to improve the accuracy of forecasting the parameters of the generating equipment of hydroelectric power plants and hydraulic structures, machine learning methods were used. The new forecast model has increased the accuracy of the forecasts by an average of 3.67%.

## 1 Introduction

The Russian electric power industry is one of the largest and most reliable energy systems in the world. The basis of the production potential of the Russian electric power industry includes more than 700 power plants with a total installed capacity of more than 230 GW and power lines of all voltage classes with a length of more than 2.5 million km.

The Unified Energy System (UES) of Russia is characterized by the presence of various types of power plants, such as Thermal Power Plants (TPP) (66% of the installed capacity of the UES), Nuclear Power Plants (NPP) (17%), Hydro Power Plants (HPP) (16%) and Renewable-energy Power Plants (less than 1%). The diversification of the energy system across three key sources of electricity generation significantly increases the efficiency and reliability of the UES of Russia compared to that of other countries. Solar and wind generation has a high growth rate, but it doesn't influence the country's UES.

Combined heat and power plants (CHP) operate in a cogeneration mode, and generate both electrical energy and heat. At these power plants electricity generation is largely dependent on the heat consumption of industrial and residential facilities located near the generating facility. Nuclear power plants operate in the basic operation mode, which is characterized by almost constant load during the whole day. Under these conditions, the non-uniformity of electric energy consumption is compensated by hydroelectric power plants, which in a short period of time (5-8 minutes) have the possibility of including the full composition of the generating equipment to the load of the installed capacities of hydroelectric power plants.

The role of hydropower plants in the Russian UES is extremely important. The following characteristics make

the hydropower plants the indispensable energy sources: high reliability of the generating equipment, lack of transport operations for fuel transportation, high maneuverability and operation speed for starting and stopping the generating equipment, the implementation of automatic secondary frequency control.

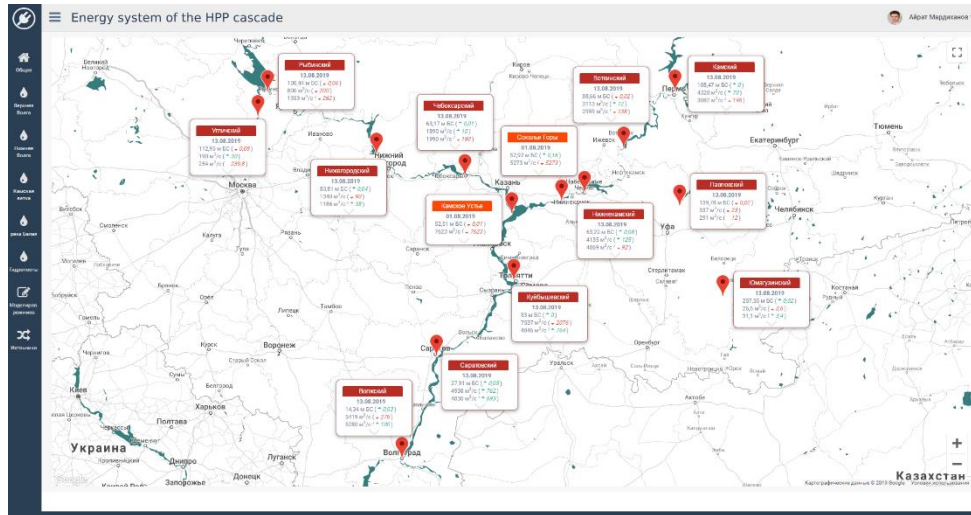
It is worth noting that in Russia the hydropower plants are mostly functioning according to a cascade scheme (Volga-Kama cascade, Angara-Yenisei cascade and others). Hydropower systems connected by a single water regime and located at the same watercourse are called a cascade of hydroelectric power plants. Construction of large cascades of hydroelectric power stations with huge reservoirs provides the most rational use of water resources at large rivers. The key feature of cascade hydropower system functioning is interconnection between the steps (system elements). First of all this means energy, hydrological and hydraulic connections.

The hydropower system functioning in the form of large cascades of hydroelectric power stations is of interest not only from the energetic point of view. It has a multi-purpose use in agricultural and fishery industries, industrial and municipal water supply, river navigation and others. A detailed description of the interests of each water user in the hydroelectric power station cascade can be found in [1].

As these interests come from different government and commercial structures, it becomes obvious that the requirements of water users in terms of establishing favorable operating conditions of the hydropower system of the HPP cascade are to a greater extent contradictory.

Currently, there is no effective model for the operation of the hydropower system of the hydroelectric power station cascade, which makes it possible to equally take into account the interests of all water users. The existing methods for optimizing the operating

\* Corresponding author: [mardikhanov@mail.ru](mailto:mardikhanov@mail.ru)



**Fig. 1.** Location of the Volga-Kama cascade of hydropower plants on the Volga, Kama and Belaya rivers in the central European part of Russia.

modes of hydropower systems either have lost their relevance or satisfy the interests of a specific circle of water users. The optimization model should take into account the interests of all users of water resources. Thus, the optimization problem has an uncertain number of criteria for optimality.

Attempts to solve similar problems were made in works [2-12]. However, for some of them it is hard to find a practical implementation, the others are used for individual hydropower plants or small cascades of hydropower plants.

## 2 Methods of finding a compromise solution

It is worth noting that the authors of the article attempted to solve the problem stated above. In particular, in article [13], the author proposed a method of successive concessions and an algorithm for calculating the operation modes of a cascade of hydroelectric stations. In the framework of the developed methodology, the problem of optimal distribution of runoff between hydroelectric power plants is reduced to a compromise satisfaction of requirements for expenditures through hydroelectric power plants or water levels in reservoirs for all water users of the cascade. Thus, the task of optimally distributing the runoff between the waterworks of the cascade in a deterministic formulation is reduced to determining the operation mode of the hydroelectric power station cascade, in which the maximum possible number of requirements are implemented ranked by importance. In addition, the specified mode restrictions are satisfied.

The developed methodology allows one to ensure the interests of energy system and at the same time satisfy the requirements of other water users and ecology.

The technique was tested using hydroelectric power plants of the Volga-Kama cascade. This cascade is one of the largest HPP cascades in the world, located at the territories of 18 regions in Russia (Figure 1).

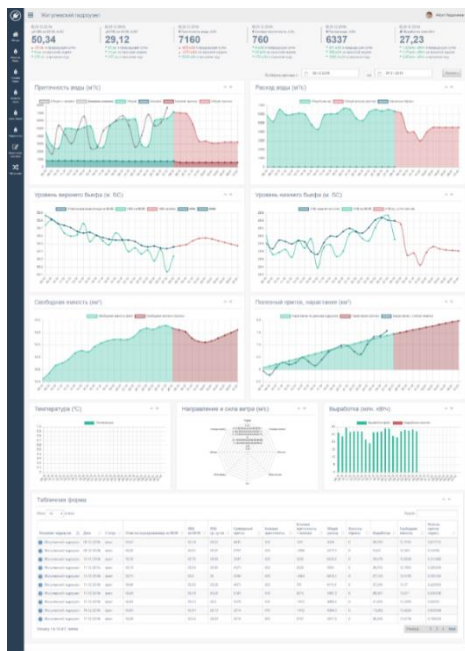
Electricity generation by hydroelectric power stations of the cascade exceeds 37 billion kWh per year. The cascade consists of 8 large hydropower plants on the Volga river: Ivankovskaya HPP, Uglichsкая HPP, Rybinskaya HPP, Nizhegorodskaya HPP, Cheboksarskaya HPP, Zhigulevskaya HPP, Saratovskaya HPP and Volgogradskaya HPP; 3 hydropower plants on the Kama river: Kamskaya HPP, Votkinskaya HPP and Nizhnekamskaya HPP; 2 hydropower plants on the Belaya river: Pavlovskaya HPP and Yumaguzinskaya HPP. The total installed capacity of the HPP cascade exceeds 11 GW (Table 1).

**Table 1.** Hydropower stations of the Volga-Kama cascade.

HPP	Capacity (MW)	Reservoir volume (km <sup>3</sup> )
Ivankovskaya	28.8	1.22
Uglichsкая	120	1.245
Rybinskaya	356.4	25.4
Nizhegorodskaya	520	8.82
Cheboksarskaya	1370	4.6
Zhigulevskaya	2467	57.3
Saratovskaya	1403	12.87
Volgogradskaya	2671	31.45
Kamskaya	552	12.2
Votkinskaya	1020	9.36
Nizhnekamskaya	1205	4.21
Pavlovskaya	166.4	1.411
Yumaguzinskaya	45	0.89
Total	11.924	170.976

It should be noted that on the basis of the developed methodology, we developed the software package “Energy system of the HPP cascade”. The software package is implemented in ASP.NET C# language using MVC technology. As a database, MSSQL Server is used, IIS (Internet Information Server) is used as a web server. Access to the software package is provided from any web browser at <http://hydrocascade.com>. The graphical

interface of the Zhigulevskaya HPP web page in the “Energy system of the HPP cascade” software complex is presented in Figure 2.



**Fig. 2.** Graphical interface of the individual page of the Zhigulevskaya HPP in the software product “Energy system of the HPP cascade”.

The developed program software is used for calculation of medium-term operation modes of the hydropower system of the Volga-Kama cascade. It is actively used by representatives of the owners of hydroelectric power stations, ministries and committees of ecology and natural resources of the regions bordering the reservoirs of the Volga-Kama cascade. The calculation results are used at meetings of the Interdepartmental Working Group on Regulation of the Operation Regimes of Volga-Kama cascade hydropower plants at the Federal Agency for Water Resources.

The software product is used by JSC "Tatenergo" to calculate the operating modes of Nizhnekamskaya HPP. Software solution has shown its economic efficiency. The actual economic effect from the use of research results for JSC "Tatenergo" (owner of the Nizhnekamskaya HPP) annually amounts to more than 100 million rubles (~ 1.503 million dollars) for one hydroelectric station.

### 3 Reducing the accuracy of forecast parameters in the developed model

After launching the “Energy system of the HPP cascade” program software, the authors continued research on improving the accuracy of forecasting the basic parameters of the generating equipment and hydraulic structures.

First of all, these studies were aimed at improving the accuracy of forecasting the downstream level of the hydropower plants. In 2012 we proposed a method for predicting the downstream level of a hydropower plant

under conditions of daily flow regulation [14]. In the framework of this methodology, a dynamic characteristic of Nizhnekamskaya HPP downstream level was obtained, which has the peculiarity of daily updating to improve the accuracy of the forecast.

It should be noted that this technique has proven itself when considering formation of short-term operation modes of individual hydropower plants. However, scaling the methodology to a cascade, which consists of 13 hydroelectric power plants, revealed a decrease in the prediction accuracy of the parameters, which in turn adversely affected the accuracy of the optimization model of the cascade hydropower system.

### 4 Application of machine learning methods

To improve the accuracy of the forecast, it was decided to apply one of the directions of artificial intelligence, i.e. machine learning. Machine learning is a set of artificial intelligence methods, the characteristic feature of which is not a direct solution of the problem, but learning in the process of applying solutions to a set of similar tasks (search for dependencies in a large amount of data). The key to successful application of machine learning methods in practical problems is the presence of a large amount of qualitative data on the object of forecasting.

At each large hydroelectric station archiving of telemetric indicators of the station takes place in real time. In particular, at the Nizhnekamskaya HPP, data for more than 10 years is stored in a database obtained in a 3-minute period. At the same time, the estimated number of parameters is more than 50. These include the performance indicators of generating equipment, hydrological and energy indicators, ambient temperature, wind strength and direction, and others. Thus, the total amount of data for each parameter is more than 1.75 million values.

For each of the 13 hydroelectric power plants there are 50 parameters, so their total number is 650. In addition, the model includes data on gauging stations that are located between hydropower plants. As a result, the total amount of training data exceeds 1 billion values. This amount of data is more than enough to train the “supervised learning” model.

#### 4.1 Supervised learning

Supervised learning is learning when there are 3 types of data, namely:

- Training set  $X_{train}$ ;
- Testing set  $X_{test}$ ;
- Predictable data  $X_{predict}$ .

$X_{train}$  is used to build/"fit" the model (choice of variables, determination of coefficients, etc.) in order to minimize the error between the predicted hydropower plant downstream level, and the actual level.  $X_{test}$  is a set for which we know the answers and at the same time we want to test our model. The prediction is formed using our model and compared with real answers. Using this

comparison, we can understand how well our model works.  $X_{\text{predict}}$  is the new data used in business (to predict the level in the downstream reach).

Thus, let  $X$  be the set of objects, and  $Y$  the set of valid answers.  $y^*$  is the objective function.  $y^*: X \rightarrow Y, y_i = y^*(x_i)$  are known only for a finite subset of the objects  $x_1 \dots x_m$  from  $X$ . In this case, the pairs  $(x_i, y_i)$  are called precedents. The set of such pairs with  $i$  from 1 to  $m$  is  $X_{\text{train}}$ .

As a method of forecasting, a generalized linear multiple regression model was applied:

$$Y = g(a + b_1 * X_1 + b_2 * X_2 + \dots + b_n * X_n) + error,$$

where *error* is the error component that cannot be calculated using predictors, and  $g()$  is a function. The inverse function to  $g$  is called a coupling function.

#### 4.2 The implementation of the predictive model in the programming language R

The forecast model was implemented in the R programming language using several specialized packages.

The mean absolute percentage error (MAPE) was chosen as the metric. There are other metrics, such as root mean squared error ([R]MSE). However, MAPE has a more intuitive interpretation and therefore this metric was chosen as the base.

The general structure of the R language code includes the following stages:

- At the first stage specialized packages and libraries are loaded via packet manager pacman. Pacman checks for library availability in the R studio development environment. If there are no libraries, automatic loading is performed. The list of loadable libraries includes: caret, ROCR, lift, glmnet, MASS, etc.

- At the second stage the source data is imported into the model. In addition, data structures are checked for categorical variables (factors) and whole and real numbers.

- At the third stage data is cleared, namely, formats, missing values, etc. For this, an additional custom function CleanData is used. It is worth noting that data omission is often not random. For categorical variables, in the absence of data, a new category is added. For continuous variables, a separate model for predicting missing data is created based on the available ones. At the same time, for each missing parameter, "surrogate" variables are added. In addition, at this stage, a combination of rare categories is carried out using the additional "Tbl" function.

- At the fourth stage the dataset is divided into test and training sets.

- At the fifth stage the model is "trained" on training data. Variables are selected using StepAIC (incremental regression). Stepwise regression is performed using the exact Akaike Information Criterion (AIC) as a criterion for including or deleting variables.

$$AIC = 2k - 2\ln(L),$$

where  $k$  is the number of parameters in the statistical model,  $L$  is the maximized value of the likelihood function of the model. The best model is recognized for which the value of AIC is minimal.

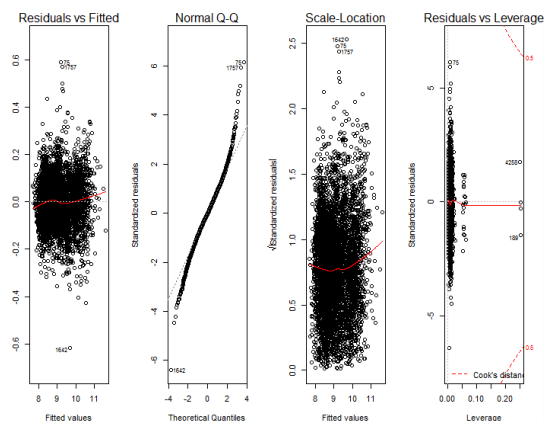
In stepwise selection, variables are added or deleted from a model one at a time, until some stopping criterion is reached. For example, in forward stepwise regression one adds predictor variables to the model one at a time, stopping when the addition of variables would no longer improve the model. In backward stepwise regression, one starts with a model that includes all predictor variables, and then deletes them one at a time until variables being removed would degrade the model quality. The stepwise regression combines the forward and backward stepwise approaches. Variables are entered one at a time, but at each step, the variables in the model are reevaluated, and those that don't contribute to the model are deleted. A predictor variable may be added and deleted from a model several times before a final solution is reached [15].

The implementation of stepwise regression methods vary by the criteria used to enter or remove variables. The stepAIC() function in the MASS package performs stepwise model selection (forward, backward, stepwise) using an exact AIC criterion [15].

- At the sixth stage we obtain a prediction model for test data.

- At the seventh stage we obtain a metric for test data.

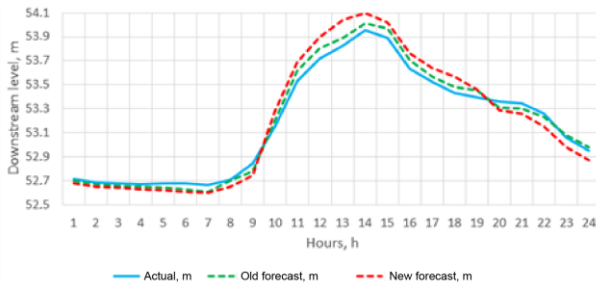
Figure 3 shows models of regression diagnosis graphs. The graphs show that systematic errors in the model have not been identified.



**Fig. 3.** Regression diagnosis graphs.

Figure 4 shows a graph comparing the actual downstream level of Nizhnenskaya HPP for June 24, 2019 with old and new forecast data. It should be noted that in the range of 52.61 - 54.01 (1.4 m), the error of the new forecast was reduced by 3.67%.

Thus, the combination of computational power methods and the availability of data open up tremendous opportunities for decision-making during formation of operation modes of hydropower systems. The use of machine learning methods can significantly increase the accuracy of basic parameters prediction for hydropower systems cascade.



**Fig. 4.** Graph of level downstream for Nizhnekamskaya HPP.

The developed software solution using the successive concessions method allows one to ensure the interests of a single energy system and at the same time satisfy the requirements of other water users and ecology.

At the next stage, it is planned to scale the forecast method to other parameters of the model and to update a new release of the considered software solution (<http://hydrocascade.com>).

The next article will be devoted to the use of the random forest method of machine learning algorithm in the model of operation mode calculation of cascade hydropower systems.

## 5 Conclusion

A technique has been developed to search for compromise solutions during formation of the operation modes of hydropower systems. Based on the developed methodology, the software solution “Energy system of the HPP cascade” (<http://hydrocascade.com>) was implemented. In the existing model, machine learning methods were used to improve the accuracy of predicting the operating parameters of hydropower equipment and hydraulic structures.

## References

1. A. Mardikhanov, Ensuring the interests of the unified energy system of Russia in the formation of the modes of operation of the hydropower complex, *Vestnik Kazanskogo gosudarstvennogo energeticheskogo universiteta*, **2(42)**, 3-11 (2019)
2. Z.K. Feng, W.J. Niu, Ch.T. Cheng, S.I. Liao, Hydropower system operation optimization by discrete differential dynamic programming based on orthogonal experiment design, *Energy*, **126**, 720-732 (2017)
3. G. Hermida, E.D. Castronuovo, On the hydropower short-term scheduling of large basins, considering nonlinear programming, stochastic inflows and heavy ecological restrictions International, *Journal of Electrical Power & Energy Systems*, **97**, 408-417 (2018)
4. M.M. Belsnes, O. Wolfgang, T. Follestad, E.K. Aasgård, Applying successive linear programming for stochastic short-term hydropower optimization, *Electric Power Systems Research*, **130**, 167-180 (2016)
5. Ch. Li, J. Zhou, Sh. Ouyang, X. Ding, L. Chen, Improved decomposition–coordination and discrete differential dynamic programming for optimization of large-scale hydropower system, *Energy Conversion and Management* volume, **84**, 363-373 (2014)
6. Y. Tan, Z. Dong, Ch. Xiong, Z. Zhong, An Optimal Allocation Model for Large Complex Water Resources System Considering Water supply and Ecological Needs, *Water*, **11(4)**, 843 (2019)
7. D.A.G. Vieira, L. Guedes, A.Ch. Lisboa, R.R. Saldanha, Formulations for hydroelectric energy production with optimality conditions, *Energy Conversion and Management*, **89**, 781- 788 (2015)
8. A. Mahor, S. Rangnekar, Short term generation scheduling of cascaded hydro electric system using novel self adaptive inertia weight PSO, *International Journal of Electrical Power & Energy Systems*, **34(1)**, 1-9 (2012)
9. J. Tian, J. Chang, Z. Zhang, Y. Wang, Influence of Three Gorges Dam on Downstream Low Flow, *International Journal of Electrical Power & Energy Systems*, *Water*, **11(1)**, 65 (2019)
10. J. Tian, J. Chang, Z. Zhang, Y. Wang, Influence of Three Gorges Dam on Downstream Low Flow, *International Journal of Electrical Power & Energy Systems*, *Water*, **11(1)**, 65 (2019)
11. F. Zhenghui, W. Yuqi, W. Lu, Z. Haojin, An inexact stochastic optimization model for multi-conflict regional water resources allocation in the South-to-North water benefited area, *Journal of Hydroinformatics*, **20(4)** (2018)
12. Ch. Cheng, H. Wu, W. Lu, X. Wu, J. Shen, J. Wang, Power Generation Scheduling for Integrated Large and Small Hydropower Plant Systems in Southwest China, *Journal of Water Resources Planning and Management*, **143(8)** (2017)
13. A. Mardikhanov, V. Sharifullin, *Search for a compromise between the interests of water users while forming the operating modes of HPP cascade*, *IOP Conf. Ser.: Mater. Sci. Eng.*, **012156** (2019)
14. A. Mardikhanov, The method of forecasting the level of the lower reach of hydroelectric power plants in the conditions of daily flow regulation, *Gidrotechnicheskoye stroitelstvo*, **6**, 16-19 (2012)
15. R.I. Kabacoff, *R in action. Data analysis and graphics with R* (2011)