Available online at www.sciencedirect.com

ScienceDirect

Energy Reports 8 (2022) 765–774

www.elsevier.com/locate/egyr

International Conference on Energy Storage Technology and Power Systems (ESPS 2022), 25–27 February 2022, Guilin, China

Medium-term forecasting of power generation by hydropower plants in isolated power systems under climate change

Murodbek Safaraliev[a], Natalya Kiryanova[b], Pavel Matrenin[c], Stepan Dmitriev[a,*], Sergey Kokin[a], Firuz Kamalov[d]

[a] *Department of Automated Electrical Systems, Ural Federal University, Yekaterinburg, Russia*
[b] *Department of Automated Electrical Power Systems, Novosibirsk State Technical University, Novosibirsk, Russia*
[c] *Department of Industrial Power Supply Systems, Novosibirsk State Technical University, Novosibirsk, Russia*
[d] *Department of Electrical Engineering, Canadian University Dubai, Dubai, United Arab Emirates*

Received 15 September 2022; accepted 27 September 2022
Available online xxxx

Abstract

Reliable operation of power systems (PS), including those with a significant share of hydropower plants (HPPs) in the energy balance, largely depends on the accuracy of forecasting power generation. The importance of power generation forecasts increases with the development of renewable power generation, which is stochastic by nature. Those kinds of tasks are complicated by the lack of reliable information on metrological data and estimated energy consumption, which is also stochastic. In the medium-term forecasting (MTF) of power generation by HPPs, the seasonality of changes in flow and inflow of water should be taken into account, which significantly affects the reserves and regulatory capabilities of the power system as a whole. This work discusses the problem of constructing a model for MTF of power generation HPP in isolated power systems (IPS), taking into account such atmospheric parameters as air temperature, wind speed and humidity. To address constant climatic changes, this paper suggests implementing machine learning models. The proposed approach is characterized by a high degree of autonomy and learning automation. The paper provides a comparative study of the machine learning models such as polynomial model with Tikhonov's regularization (LR), k-nearest neighbors (kNN), multilayer perceptron (MLP), ensembles of decision trees, adaptive boosting of linear models (ABLR), etc. Computational experiments have shown that the machine learning approach yields the results of sufficient quality, which allows to use them for forecasting of power generation HPP in isolated power systems under conditions of climate change. The Adaptive Boosting Linear Regression model is the simplest and most reliable machine learning model that has proven itself well in the tasks with a relatively small amount of training samples.

Peer-review under responsibility of the scientific committee of the International Conference on Energy Engineering and Power Systems, EEPS, 2022.

Keywords: Medium-term forecasting of power generation; Hydropower plant; Isolated power system; GBAO; Ensemble models; Climate change; Temperature

* Corresponding author.
E-mail address: sa.dmitriev@urfu.ru (S. Dmitriev).

https://doi.org/10.1016/j.egyr.2022.09.164

Peer-review under responsibility of the scientific committee of the International Conference on Energy Engineering and Power Systems, EEPS, 2022.

1. Introduction

Hydropower today remains a significant and no less environmentally friendly type of renewable energy source (RES) for electricity generation worldwide, providing 19% of the electricity of the entire globe [1,2]. In most cases, a small hydroelectric power plant (SHPP) will be built and operated on a river without a dam or a small reservoir and represents the most cost-effective and environmentally friendly technologies for electricity production, which should be considered for implementation to provide electricity to decentralized areas in less economically developed countries [3–6].

Accurate and reliable forecasting of power generation by the HPPs is essential for the management and planning of hydropower generation. Due to the rapid development of distributed generation, there is not enough historical data on power generation, which makes it difficult to develop forecast models.

The complexity of forecasting the generation of HPPs is because of several reasons: complex nonlinear dependencies of hydrological parameters; influence of various meteorological factors; the influence of factors such as the level of snow in the mountains or the amount of precipitation in previous months extended over time; changes in the dependencies because of changes in climate and topography; the need to use long-term observations (large datasets) [7–9].

The task of forecasting of power generation HPP is considered to be one of the most difficult tasks in the field of intellectual data analysis; it requires complex analysis of large amounts of data with view to the influence that multiple relationships and dynamic processes have on power consumption.

MTF of power generation for each type of RES is a key issue for a PS, since such forecasting is the basic tool for ensuring reliable power supply, organizing maintenance, planning power reserves, and repairs, and maintaining the regime. MTF is critical for the owners of the renewable energy resources in order to define their behavior in the electricity market, schedule maintenance tasks and set targets with a one-week horizon.

Most of the published papers dedicated to the development of MTF models and related to hydropower issues are focused on management of water resources based on projection data for the water inflow to reservoirs only [10–12], or take into account the information on the water runoff and inflow [13,14]. A number of studies are aimed at assessing the capacity and planning of generation for mini-HHPs [15], small-HHPs [16], and large HPPs [17]. As a rule, the models applied for medium-term (one week ahead) forecasting of power generation do not take into account climatic changes. Climate change usually occurs gradually, but unpredictable short-term changes can arise. As a result, machine learning models built using meteorological factors with good results during testing can significantly decrease forecasting accuracy after some time due to changing external conditions. The risk of such a scenario is especially high in the absence of data from a long history of observations. Therefore, it is necessary, firstly, to analyze the influence on the model accuracy even of those factors, the benefits of which seem obvious. Secondly, it is reasonable to apply models that are not prone to overfitting and can be re-trained on a relatively small dataset when climatic conditions change, adapted to the changes. There are a lot of medium-term forecasting models based on different nature: autoregression models, ensemble regression trees, neural networks (from compact multilayer perceptrons to deep recurrent networks). Besides choosing a model, it is necessary to analyze the initial data pre-process the data, as well as select the characteristics.

This paper studies the implementation of compact models based on machine learning (ML) for medium-term forecasting of power generation HPP under climate change. The characteristic feature of applied ML models is relatively small model size. It allows to train and tune models quickly in the automation model of training and eliminate the risk of overtraining.

2. Methodology

2.1. Assessment of energy resources of the research object

The object of the study is the IPS of the Gorno-Badakhshan Autonomous Oblast (GBAO) — a region located in the eastern part of the Republic of Tajikistan (RT) and Central Asia. GBAO is one of the richest regions of the RT in terms of reserves and potential of hydropower, but today less than 1 percent of the existing potential is used. The hydropower resources of small and large mountain rivers in the region are so large that when the level of their use reaches the range of 20%–25%, GBAO will return to one of the richest regions of the country. The existing hydropower potential of GBAO reflects the economic efficiency and commercial benefits to justify its use, as well as for the construction of small, medium and large HPPs [18,19].

To date, the network of hydrometeorological observations in the RT and in particular in the studied region is considered less developed and studied. Thus, it will be difficult to assess the real potential of wind and solar energy in the region [20,21].

Fig. 1 shows the geographical location of the HPP and substations in the GBAO IPS for illustrative representation.

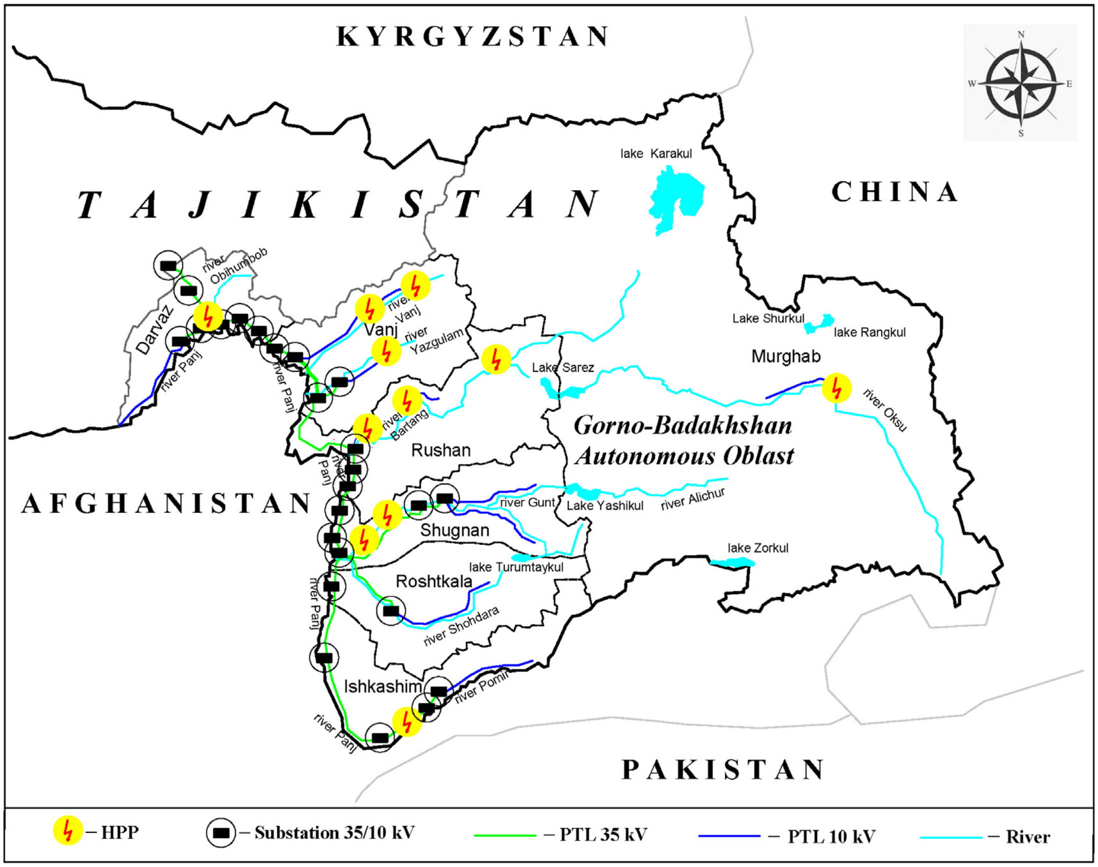

Fig. 1. Geographical location of the Gorno-Badakhshan Autonomous Oblast electric power system.

The IPS of GBAO is operated by "Pamir Energy" company, the power of which reaches from 500 kW to tens of MW. For a decade, the IPS GBAO has been experiencing difficulties due to the lack of sustainable power supply during the heating season, starting from the end of October and until the beginning of April. The main reasons for the shortage of electricity are; water shortages in winter, lack of electricity market, isolated operation of the PS, lack of large storage and increased consumption during this period [22–24].

2.2. Problem statement and initial data

The following forecasting task was formulated:

$$Y^* = f(X) \tag{1}$$

where Y^* is the week ahead forecast of daily power generation, X is the input data (features) vector, $f(X)$ is a predictive model:

As the main quality indicator mean absolute percentage error (MAPE) was used:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^* - y_i}{y_i}\right| \tag{2}$$

where n – is the dataset size (number of days); $y^*{}_i$ – predicted value for the ith day; y_i – actual value.

The applied dataset contains daily values of generated power (in the isolated power system of GBAO) from November through March for the 5 years (2015–2019), and average daily temperatures, data for 765 days. For data processing, the numbers of the months were converted as follows: November = 1, December = 2,... March = 5. Year count starts from 2015, so the year numbers are converted to the range 1–5. Fig. 2 shows generation graphs for the same month (January).

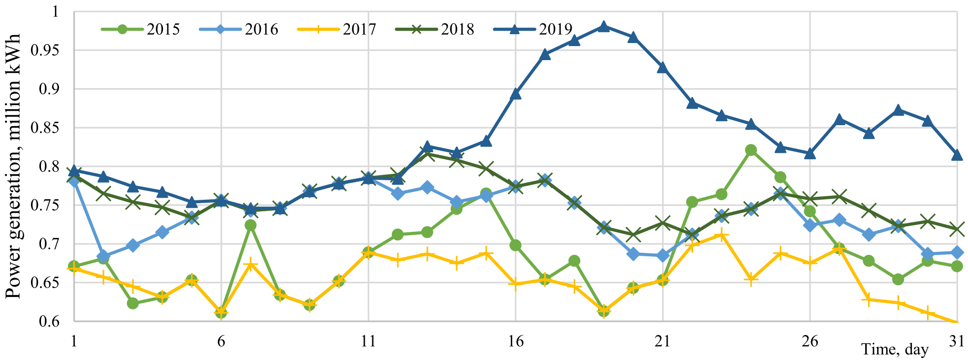

Fig. 2. Fragment of the data sample (January 2015–2019).

2.3. Data analysis and sampling for applying machine learning models

The first stage of the creation forecasting model is feature extraction. The calendar features, previous power values were used. This research also checks the hypothesis about the influence of the average daily temperature on forecast accuracy. Fig. 3 shows a matrix of Pearson's correlation coefficients for factors that can be used in forecasting. In addition, the retrospective data on power generation during the previous days is used. One can see that out of meteorological factors, temperature has significant influence (coefficient 0.357); besides, there is a noticeable year on year change (coefficient 0.375).

The resulting structure of the data set for building regression models is shown in Table 1, where m – is the length of the used time series of historical values. If a week is used, from $m = 7$, two weeks – 14, etc. For example, if $m = 7$ and a forecast is given for January 26, then the projection will use the retrospective data from January 19 ($G7$, 26 – 7 = 19) through 13 ($G13$, 26 – [7 + m – 1] = 13).

Table 1. Sample structure for applying ML models and an example of filling.

Day	Month	Year	Average daily temperature, C0	G7, mil kWh	G8, mil kWh	...	Gm + 7 − 2, mil kWh	Gm + 7 − 1, mil kWh	G, mil kWh (forecast target)
26	3	2	−3.2	0.687	0.721	...	0.684	0.782	0.724

The number of samples in the dataset depends on m, since there is no sufficient retrospective data for the first $m + 7$ days of the initial month (November). The dataset is divided into training–validation and test sets in a ratio of 4 to 1 (in this case, the last year of the whole dataset falls into the test set). It should be noted that the dataset was not shuffled; as a result, the model is trained on the retrospective data and tested on the newest data. It brings the testing process closer to real-life conditions. Table 2 shows the dataset sizes for different values of m.

Table 2. Sets sizes for different intervals of historical data used.

m	Number of columns	Size of training set (days)	Size of training set (days)	Size of training set (days)
7	11	576	110	137
14	18	547	104	130
21	25	518	98	123

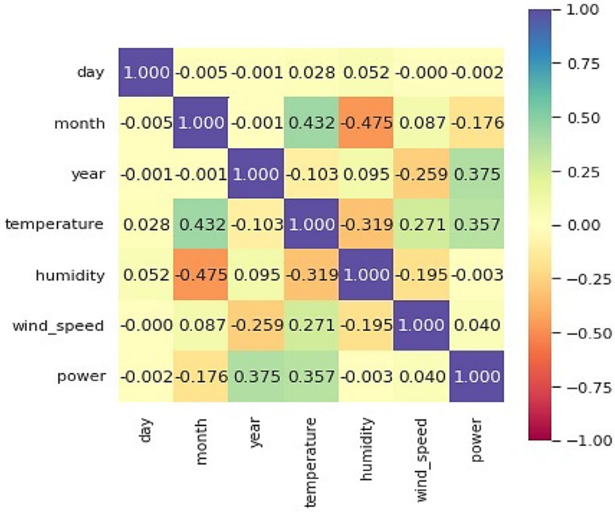

Fig. 3. Matrix of correlation coefficients.

2.4. Machine learning models applied

The ML models, the selected hyper-parameters and ranges of their values are shown in Table 3. The Random Search algorithm was applied to tune the models' hyper-parameters. The selection of hyper-parameters was performed with *m* equal to 14. Software implementations of the models, except for XGBoost [25], were taken from the open library Scikit-Learn [26]. For hyper-parameters not listed in Table 3, the default values from the specified sources [25,26] are used. The Random Search algorithm was applied to tune the models' hyper-parameters.

For all models, we analyzed the influence of the length of the time series of previous generation values and the influence of temperature on the accuracy of forecasts.

Table 3. Hyper-parameters of the applied machine learning models.

Model	Hyper parameter	Max. meaning	Max. meaning	Step	Matched value
Linear/polynomial regression with Tikhonov regularization [4*] (LR)	Polynomial degree	1	3	1	1
k- nearest neighbors (kNN)	Number of nearest neighbors k	2	8	1	2
Adaptive Boosting Decision Trees (ABDT)	Number of base models Depth of trees Minimum data points for branching	10 2 2	50 6 8	1 1 1	14 3 3
Adaptive Boosting Linear Models (ABLR)	Number of base models	2	10	1	2
Random Forest (RF)	Number of base models Depth of trees Minimum data points for branching	10 2 2	50 10 8	1 1 1	37 5 2
Extreme Gradient Boosting (XGB)	Number of base models Depth of trees Number of base models	10 2 0.01	50 10 0.5	1 1 0.05	29 4 0.3
Multilayer Perceptron (MLP)	1th hidden layer neurons 2th hidden layer neurons	10 10	200 100	10 10	90 80

3. Results and discussion

3.1. Results of machine learning models

The results of implementation of the models for various variants of the used input features on the test set are shown in Tables 4–5 and Fig. 4.

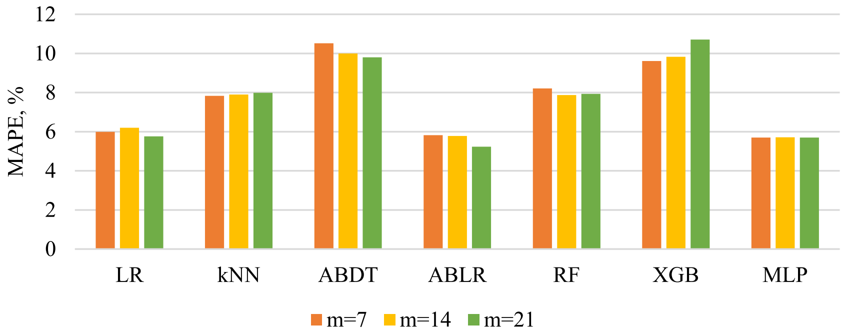

Fig. 4. Comparison of model results.

Table 4. Comparison of the results of the models (MAPE, %); in each line, the best results obtained for the corresponding model are shown in bold type.

Model	**Without using temperature**			**Using temperature**		
m	**7**	**14**	**21**	**7**	**14**	**21**
LR	5.99	6.20	**5.76**	6.53	7.15	7.19
kNN	**7.83**	7.90	7.99	13.41	14.12	13.35
ABDT	10.52	10.00	**9.80**	10.78	10.35	10.01
ABLR	5.82	5.78	**5.23**	6.25	7.30	6.98
RF	8.21	7.87	7.93	8.29	8.13	**7.72**
XGB	**9.61**	9.83	10.71	10.84	11.80	12.76
MLP	**5.70**	5.71	**5.70**	8.69	9.86	10.12

Table 5. Comparison of the results of the models (mean absolute err, MWh); in each line, the best results obtained for the corresponding model are shown in bold type.

Model	**Without using temperature**			**Using temperature**		
m	61	122	183	61	122	183
LR	44	52	44	52	61	61
kNN	61	61	61	113	122	113
ABDT	87	87	78	87	87	87
ABLR	44	44	44	52	61	52
RF	70	61	61	70	70	61
XGB	78	78	87	87	96	104
MLP	44	44	44	70	78	87

3.2. Discussion of results

The obtained results lead to the following conclusions.

1. Ensembles of decision trees (ABDT, RF, XGB), that is, models with discrete (piecewise constant) output, are inferior to models with continuous output by 2–5 percentage points.

2. For models using linear regression (LR and ABLR), an increase in the interval of the applied historical data slightly reduces the error, possibly due to a more accurate computation of trend.

3. The best accuracy was obtained using adaptive boosting with linear regression (ABLR) as the base model, which coincides with the results obtained earlier for the projection of the power consumption of this system [27].

4. The ABLR has an ultra-small model size compared to an ensemble of decision trees or even a small neural network. As a result, it significantly reduces the risk of overfitting. Also, the learning speed of this model is very high. Consequently, the ABLR model can be retrained on new data automatically and autonomously without the need for control by a specialist who would analyze the model's results and adjust its hyper-parameters.

5. A significant decrease in the accuracy of models when using meteorological data is a somewhat unexpected result that requires additional analysis. Probably, climate change from year to year also leads to a change in the dependencies between power generation and temperature, so that the model trained on the data of the first four years detects the dependencies between temperature and generation that turn out to be different the following year.

This is confirmed by the visualization shown in Figs. 5–7. The analysis shows that the dependency of generation on temperature varies significantly from year to year. Even the correlation coefficients differ significantly, in 2015 the Pearson correlation coefficient between generation and temperature being 0.38, while in 2019 it amounted to 0.77, that is, in 2019, the temperature had a significantly greater effect on generation. Therefore, when a model is trained on historical data, it can detect certain dependencies that discontinue to work in the future due to the reasons above.

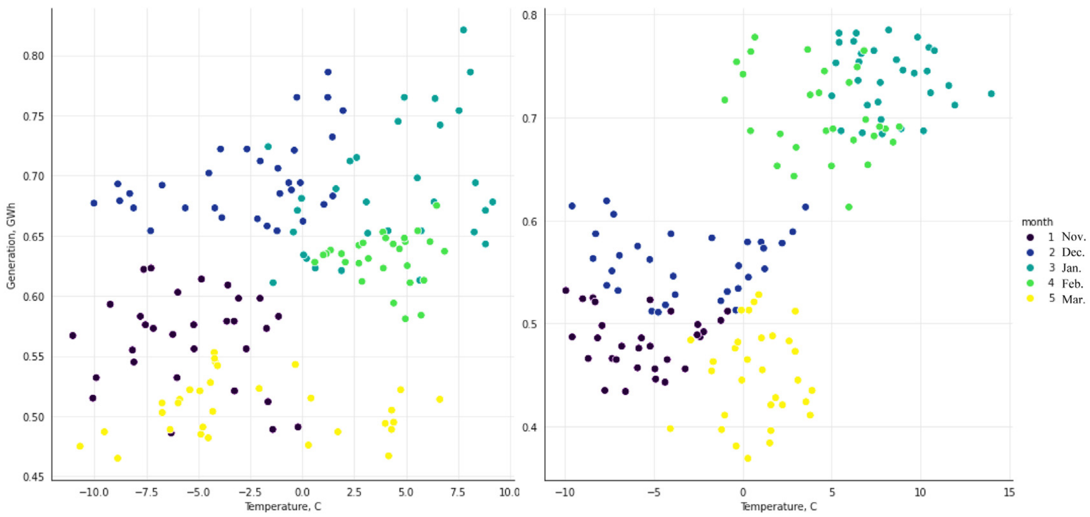

Fig. 5. Distribution of days by generation and temperature, 2015 on the left, 2016 on the right.

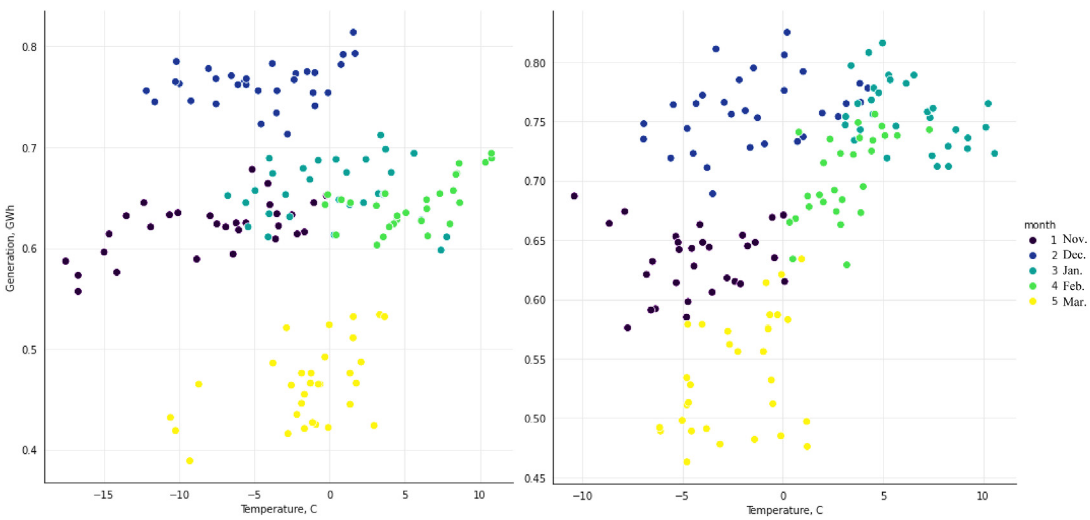

Fig. 6. Distribution of days by generation and temperature, 2017 on the left, 2018 on the right.

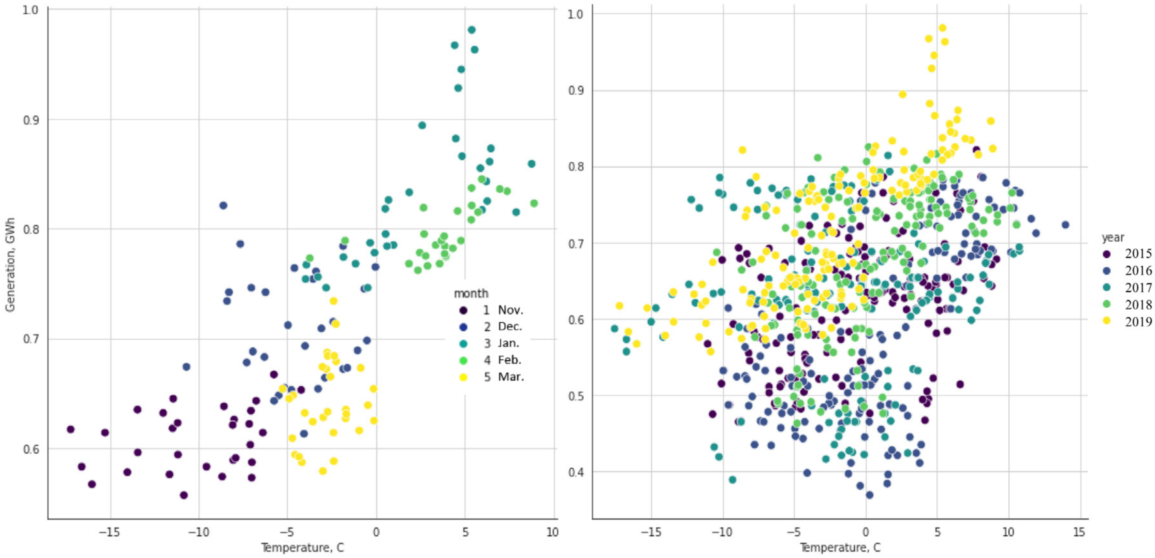

Fig. 7. Distribution of days by generation and temperature, 2019 on the left, 2015–2019. on the right.

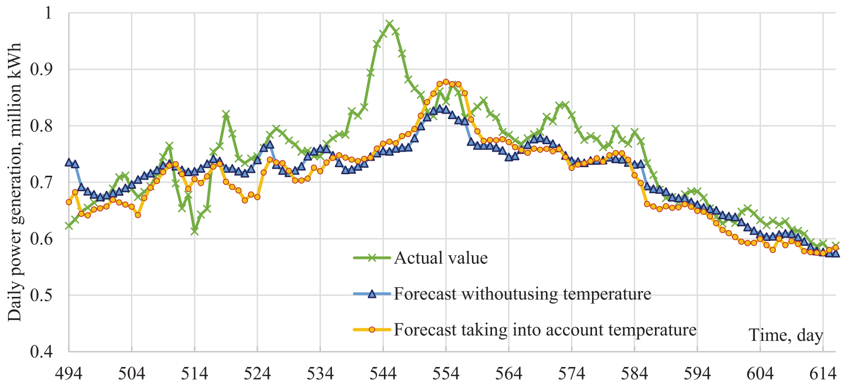

Fig. 8. Comparison of forecasts and actual power values, 2019 (test set).

The right part of Fig. 7 also shows that the year of 2019, which was included in the test set, is generally characterized by a higher generation value than the previous 4 years, especially at temperatures above 0 C^0. Comparison of the forecast obtained by the best suitable model for this task (Adaptive boosting of linear regressions with Tikhonov's regularization) is shown in Fig. 8. The average error in the test sample was 5.23%, or 44 MWh with an average daily generation of 718 MWh. Tikhonov's regularization (L2-regularization) can be written down as follows:

$$w^* = \arg\min_{w} \left(\sum_{i=1}^{n} (y_i - (wX_i - b))^2 + \lambda \left\| w \right\|_2^2 \right) \tag{3}$$

In expression (3) w, b are regression coefficients, $\|w\|_2{}^2$ – the element of the loss-function to reduce the risk of overfitting.

Since the adaptive boosting model is a linear sum of the base regressors, after the training is completed, the final ensemble model combining 4 linear models can be folded into a single one.

$$y_i^* = f(x_i) = a_1 year_i + a_2 month_i + a_3 day_i + a_4 E_{7,i} + a_5 E_{8,i} + \cdots a_{25} E_{28,i} + b$$

4. Conclusion

In this paper, the main attention is given to the development of a model for MTF of generation HPP power for a week ahead in IPS with a high proportion of HPPs. Seven different models were analyzed to forecast power generation of HPPs based on retrospective generation data and meteorological parameters. The best result was obtained using adaptive boosting with linear regression as the base model.

The obtained results allow us to conclude with a high degree of accuracy that compact ensemble ML models are acceptable for the HPP medium-term power generation prediction under temperature changes. The proposed forecasting method allows to accurately estimate the projected power reserves and opens up the possibility for optimization of power generation with a view to climate change.

The results obtained in the course of this study can be used to improve the forecasting of power generation by the IPS of GBAO, which mainly operates at the expense of HPPs, when making informed decisions about the structure of power generation in the region. The proposed methods can also be used for other power supply companies operating IPS.

Next line of efforts involves the creation, study and verification of an adaptive model with the possibility of periodic additional training in line with apparition of new data and replacement of the outdated part of the training sample. Such model would constantly and regularly adapt to changing operating conditions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

The study was financially supported as part of the Novosibirsk State Technical University, Russia development program, scientific project C22-15.

References

[1] H.S. Sachdev, A.K. Akella, N. Kumar, Analysis and evaluation of small hydropower plants: A bibliographical survey, Renew Sustain Energy Rev (51) (2015) 1013–1022.
[2] T.S. Kishore, E.R. Patro, V.S. Harish, A.T. Haghighi, A comprehensive study on the recent progress and trends in development of small hydropower projects, Energies 14 (10) (2021) 2882.
[3] M.S. Asanov, et al., Algorithm for calculation and selection of micro hydropower plant taking into account hydrological parameters of small watercourses mountain rivers of central Asia, Int J Hydrog Energy 46 (75) (2021) 37109–37119.
[4] Y. Bayazıt, R. Bakış, C. Koç, A study on transformation of multi-purpose dams into pumped storage hydroelectric power plants by using GIS model, Int J Green Energy 18 (3) (2021) 308–318.
[5] A. Ghulomzoda, et al., Recloser-based decentralized control of the grid with distributed generation in the lahsh district of the rasht grid in tajikistan, Cent Asia Energies 13 (14) (2020) 3673.
[6] A.M. Mayeda, A.D. Boyd, Factors influencing public perceptions of hydropower projects: A systematic literature review, Renew Sustain Energy Rev 121 (2020) 109713.
[7] S. Ghimire, Z.M. Yaseen, A.A. Farooque, et al., Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks, Sci Rep 11 (2021) 17497.
[8] C. Xu, L. Xiong, V.P. Singh, Black-box hydrological models, in: Handbook of hydrometeorological ensemble forecasting, Springer, Berlin, 2017, pp. 1–48.
[9] J.P. Nair, M.S. Vijaya, Predictive models for river water quality using machine learning and big data techniques - A survey, in: 2021 International conference on artificial intelligence and smart systems, ICAIS, 2021, pp. 1747–1753.
[10] Shichao Xu, et al., Baipenzhu reservoir inflow flood forecasting based on a distributed hydrological model, Water 13 (3) (2021) 272.
[11] M.E. Banihabib, R. Bandari, M. Valipour, Improving daily peak flow forecasts using hybrid Fourier-series autoregressive integrated moving average and recurrent artificial neural network models, AI 1 (2) (2020) 263–275.
[12] H. Albo-Salih, L. Mays, Testing of an optimization-simulation model for real-time flood operation of river-reservoir systems, Water 13 (9) (2021) 1207.
[13] P. Sharma, D. Machiwal, Streamflow forecasting: overview of advances in data-driven techniques, Adv Streamflow Forecast (2021) 1–50.

[14] O. Oyebode, D. Stretch, Neural network modeling of hydrological systems: A review of implementation techniques, Nat Resour Modeling 32 (1) (2019) e12189.

[15] R. Peña, A. Medina, O. Anaya-Lara, J.R. McDonald, Capacity estimation of a minihydro plant based on time series forecasting, Renew Energy 34 (5) (2009) 1204–1209.

[16] G. Li, C.X. Liu, S.L. Liao, C.T. Cheng, Applying a correlation analysis method to long-term forecasting of power production at small hydropower plants, Water 7 (9) (2015) 4806–4820.

[17] W. Zhong, J. Guo, L. Chen, J. Zhou, J. Zhang, D. Wang, Future hydropower generation prediction of large-scale reservoirs in the upper Yangtze River basin under climate change, J Hydrol 588 (2020) 125013.

[18] A. Ghulomzoda, et al., A novel approach of synchronization of microgrid with a power system of limited capacity, Sustainability 13 (24) (2021) 13975.

[19] V. Manusov, et al., Optimal management of energy consumption in an autonomous power system considering alternative energy sources, Mathematics 10 (2) (2022) 1518194.

[20] M.K. Safaraliev, I.N. Odinaev, J.S. Ahyoev, K.N. Rasulzoda, R.A. Otashbekov, Energy potential estimation of the region's solar radiation using a solar tracker, Appl Solar Energy 56 (4) (2020) 270–275.

[21] M. Safaraliev, et al., The transient analysis of the hydrogenerator of Nurek HPP subject to automatic excitation control action, Prz Elektrotech 8 (2020) 35–38.

[22] P. Matrenin, et al., Adaptive ensemble models for medium-term forecasting of water inflow when planning electricity generation under climate change, Energy Rep 8 (1) (2022) 439–447.

[23] A.K. Kirgizov, et al., Expert system application for reactive power compensation in isolated electric power systems, Int J Electr Comput Eng 11 (5) (2021) 3682–3691.

[24] Sultonov Sh, et al., Specifics of hydropower plant management in isolated power systems, Prz Elektrotech 4 (2022) 53–57.

[25] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, 2016, pp. 1–13, [Web source]. Available online: https://arxiv.org/abs/1603.02754.

[26] Machine learning in python, 2021, [Web source]. Available online: https://scikit-learn.org.

[27] P. Matrenin, et al., Medium-term load forecasting in isolated power systems based on ensemble machine learning models, Energy Rep 8 (1) (2022) 612–618.