



Forecasting the hydroelectric power generation of GCMs using machine learning techniques and deep learning (Almus Dam, Turkey)

Hesham Al Rayess and Asli Ülke Keskin

Ondokuz Mayıs University, Engineering Faculty, Civil Engineering Department, Samsun, Turkey

Received 21 October 2020, in final form 14 April 2021

Renewable energy is one of the most important factors for developed and sustainable societies. However, its utilization in electrical power grid systems can be very challenging regarding rates predictably. Renewable energy depends mainly on environmental conditions such as rainfall-runoff ratios and temperature. Because of that, the expected power production heavily fluctuates, which makes the prediction and calculation of feed-in into the power grid very challenging. The accurate forecasting of energy production is a very crucial issue for power management process. This paper presents the results of deploying Machine Learning Techniques in short-term forecasting of the amount of energy produced of General Circulation Models (GCMs) Data by Almus Dam and Hydroelectric Power Plant in Tokat, Turkey. The study demonstrates the use of modeling techniques in hydropower forecasting process using the predicted monthly hydroelectric power generation data of GCMs from 2018 to 2080. Decision Tree, Deep Learning, Generalized Linear, Gradient Boosted Trees and Random Forest models are utilized to forecast the hydropower production. The results show that the correlation value of the gradient boosted trees model equals 0.717, which means that the gradient boosted trees model is the most successful model for the present data. The gradient boosted trees model used in the prediction process for each GCM in each scenario is 4.5 and 8.5. The results show that there are small differences between the models, which means that the predictions are going in similar directions for all these models.

Keywords: renewable energy, hydropower, machine learning techniques, deep learning, general circulation model, Turkey

1. Introduction

Energy plays an important role in the economic and social development and in improving living standards. Hydropower is one of the most important sources of energy in the globe. Turkey is one of the rapidly growing countries in the world with constant growth of economy and population. This means that there is a

constant increase in Turkey's demand of energy. Turkey has many renewable energy resources that can essentially supply all the country energy demands (Yüksel, 2008).

Hydroelectric energy is a very clean and renewable energy source produced from electrical generators, which convert motion into energy as water falls in dams. Hydropower represents more than 92.0% of the electricity generated from renewable resources worldwide (Smith et al., 2012). Hydroelectric energy has zero emissions and low running cost (Gökgöz and Filiz, 2018). Moreover, hydroelectric power is preferred by many countries for its technical, economic, and environmental benefits (Huang and Yan, 2009).

In a hydroelectric power plant (HEPP), turbines convert water pressure into mechanical power. The produced power is proportional to the product of water discharge and pressure head (Yüksek et al., 2006). Turkey is considered as a recently developed country and an economic center of the region. As a result of its fast-growing economy coupled with its increased young population, the demand of electricity in Turkey is continuously increasing. Turkey is classified as one of the net energy importer countries (Melikoğlu, 2013a). Turkey imports most of its energy sources like petroleum, coal, and natural gas which affects negatively on the macroeconomic balance within the country (Melikoğlu, 2013b).

In the second half of the 20th century, Turkey started to focus on the construction of dams and subsequently hydropower plants. These dams were also used for supplying water and irrigation purposes. Hydropower is one of the most important renewable energy sources in Turkey. Moreover, hydropower is the second largest domestic energy source after coal in Turkey (Kaygusuz, 2002).

In Turkey, the gross annual hydroelectric potential was estimated to be about 433 GWh/year. The number of hydropower plants under operation has reached more than 596 with an installed capacity of approximately 26.8 MW and the annual average generation is approximately 93.7 GWh at the end of 2016 (TEDAS, 2016).

Because of rapid technology development, Data mining became one of the most effective methods in processing data sets. Machine learning (ML) is a subset branch of artificial intelligence. ML are specific computer algorithm techniques which are developed through trials and experience (Awad and Khanna, 2015). ML algorithms can be used in building models based on a sample of data to make a prediction of decisions without any direct programming. ML has been previously used in hydropower prediction and water resources management (Mosaavi et al., 2019).

The energy power forecasting studies in Turkey have started in the 1960s, and as of 1984, econometric models have been applied for forecasting purposes (Kankal et al., 2011).

Several studies have examined the impact of climate change on hydropower plants all over the world. Forrest et al. (2018) used projections from four climate

models under Representative Concentration Pathways (RCP4.5 and RCP8.5) to evaluate the impact of climate change conditions on California hydropower generation. This study made an evaluation for one small hydropower without making any forecast to the energy production. Beyene, Lettenmaier, and Kabat (2010) predict increasing precipitation and temperature in Nile River Basin with an overall positive effect on production from the Aswan Dam. However, it is expected that this effect will turn into an overall negative effect toward the end of the century.

In this paper, the main objective is to forecast the hydroelectric power generation (the energy production) of GCMs using machine learning and deep learning algorithms in Almus Hydroelectric Power Plant for the interval between 2018 to 2080. These algorithms are Decision Tree (DT), Deep Learning (DL), Generalized Linear (GL), Gradient boosted trees (GBT) and Random forest (RF). This historical data-based forecasting examines future conditions of energy production and the effectiveness of using these models to understand the behavior of the system.

2. Study area

Almus Dam is an earthen embankment dam that is near the town of Almus. It is located 28 kilometers from Tokat city in the north east of Turkey ($40^{\circ} 24' 27''$ N, $36^{\circ} 54' 11''$ E). The dam is constructed on the Yesilirmak River, which runs into the Black Sea. Almus Dam was mainly constructed for irrigation, flood control and hydroelectricity purposes. The hydroelectricity power plant, established in 1966, has a capacity of 27 megawatts. The plant is divided into three facilities;

Table 1. Almus Dam and HEPP characteristics (General Directorate of State Hydraulic Works, 2018).

Dam name	Almus
Province	Tokat
River	Yesilirmak
Minimum operating level (altitude)	767.37 meters
Maximum operating level (altitude)	804.5 meters
Water volume of minimum operating level	151,473 thousand m ³
Water volume of maximum operating level	1,006,730 thousand m ³
Water used for power generation	855,257 thousand m ³
Lake area	31 km ²
Height from the riverbed	78 meters
Irrigation area	21,350 ha
Power	27 MW
Annual production	99 Gwh



Figure 1. The location of the Almus Dam and HEPP in Turkey.

Table 2. Part of the data input of Almus HEPP for the model.

Month	Production (KWh)	Total incoming water (million m ³)	Incoming flow m ³ /s
January-71	8,410,700	32.1	12.0
February-71	9,522,414	20.81	8.6
–	–	–	–
–	–	–	–
December-13	2,977,680	14.3	5.32

each facility has a 9-megawatt capacity (General Directorate of State Hydraulic Works, 2018). Figure 1 shows the location of the Almus Dam and HEPP in the eastern north part of Turkey (Tokat). Table 1 shows the characteristics of the Almus Dam and HEPP in meter and cubic meter units for levels and volumes, respectively.

The data of the Almus hydropower used in the study are shown in Tab. 2. The data consist of monthly data of energy production of the dam, the total incoming water as a volume, incoming water as discharge and the lake water level as a depth. The data were collected from the electricity generation company (EUAS) and The Turkish State Meteorological Service. Monthly data is available up until 2013, but from the interval 2013 to 2021, the only available data is Annual.

3. Methods

In this study, RapidMiner Studio program was used for estimating the most accurate model. This program works using the most common models (Decision Tree (DT), Deep Learning (DL), Generalized Linear (GL), Gradient boosted trees (GBT) and Random forest (RF)). The software uses R language for applying the machine learning algorithm. According to the correlation and Absolute error values, the predicted energy productions were estimated using the most accurate model. The functions of the model are already built in the software with some edits.

3.1. Decision tree model

Decision Tree (DT) is a type of non-parametric learning method for classifying and regression. The major purpose of this method is to start a model or pattern. This pattern will forecast the value of the aim variable by getting a new technique and simple decision rules resulted from the data features. The decision tree method is a calculation and estimation model and is used for data collection. An upside-down tree is created from general to specific during the model training process (Sattari et al., 2012).

3.2. Deep learning model

Deep Learning (DL) techniques are classified as a significant part of Machine learning (ML) methods based on Artificial Neural Network (ANN). DL techniques have been applied in analyzing, calculating, processing and detection tasks. DL is the developing output of ANN which makes forecasting more accurate and gives positive results for studies of long-term periods. DL can multiply layers of models to learn and support representation of data. Deep neural networks increase the performance of load forecasting by making a focus on parameter optimization (Gökgöz and Filiz, 2018).

3.3. Generalized linear model

The generalized linear model (GLM) is a very easy and flexible generalization of ordinary linear regression that allows for the variables of response. The GLM popularizes linear regression by permitting the linear model to be linked to the response variable through a link function and by permitting the magnitude of the difference or variance of each measurement to be a function of its forecasted value (Nelder and Wedderburn, 1972).

3.4. Gradient boosted trees model

Gradient Boosting Tree (GBT) is one of the most powerful techniques for building predictive models. The idea of this technique initiated in the observations that boosting can be explained as a special algorithm (Optimization algorithm). This observation was made by Breiman (1997). Friedman (1999) developed this explicit regression gradient boosting algorithms.

3.5. Random forest model

Random Forests (RFs) is a very new ML technique that is very powerful and easy ensemble classifiers (Krušić et al., 2017). Breiman (2001) studied and developed this type of model based on decision trees. RF generate a multiple tree according to random bootstrapped samples of the training dataset (Breiman, 2001). The concept of the technique is to run a random binary tree which uses a subset of the observations through bootstrapping techniques. The model is built from the data sampled from the original data (Catani et al., 2013). The RF forecasts the significance of variables by searching for the amount of forecasting error increases (Liaw and Wiener, 2002; Catani et al., 2013).

4. Application and results

In this study, DT, DL, GL, GBT and RF are utilized to forecast the hydro-power production of Almus Dam. Five hundred and sixteen monthly data samples were used in this paper from Almus Dam data (Production, Total incoming water, and Incoming flow) to operate the three models. Figure 2 shows the absolute error values between the applied models for the historical data of Almus HEPP. Figure 3 shows the correlation values of the applied models for the historical data of Almus HEPP.

Figure 2 makes the analysis easier to get the most accurate model between all ML models. It shows that the gradient boosted trees model has the lowest value in comparison to other models. This means that GBT model is the most accurate model to apply in forecasting the energy production with GCMs data.

In Fig. 3, the results show that the correlation value of the gradient boosted trees model equals 0.717, which means that the gradient boosted trees model is the most successful model for these data. As a result of the correlation value of the gradient boosted trees model, the model is used in the prediction process for each General Circulation Model (GCM). Figures 4 to 8 show the scatter plot between the historical data of production and prediction of production for each model.

Figures 9 to 11 show the prediction production of electricity using the gradient boosted trees model for three types of GCM in each scenario (4.5 and 8.5).

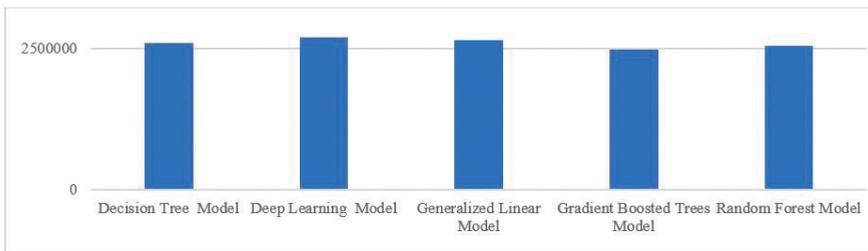


Figure 2. The Absolute Error values between the applied models for the historical data of Almus HEPP.

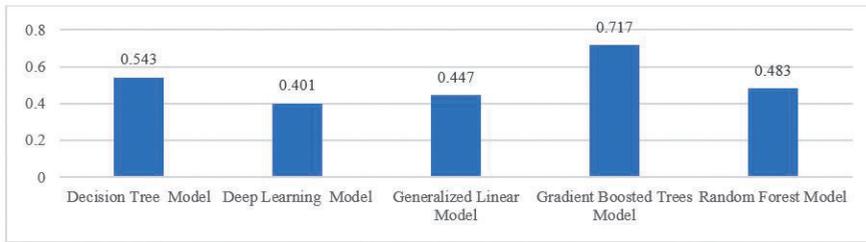


Figure 3. The correlation values of the applied models for the historical data of Almus HEPP.

The prediction of the energy production using GFDL model with 4.5 and 8.5 scenarios are shown in Fig. 9. Despite of precipitation decrease and temperature increase of GFDL 8.5, the results of GFDL 8.5 show high values of energy pro-

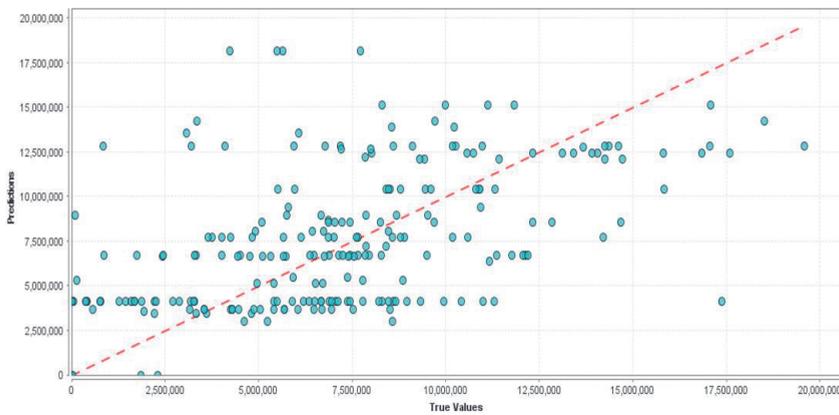


Figure 4. The scatter plot between production and prediction of production of Decision Tree Model.

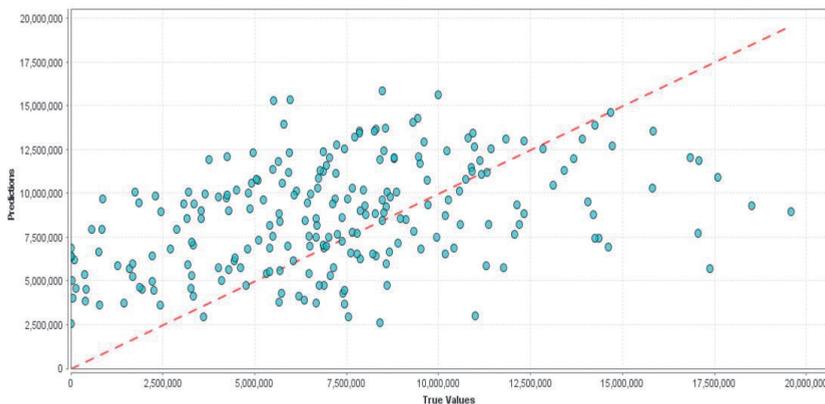


Figure 5. The scatter plot between production and prediction of production of Deep Learning Model.

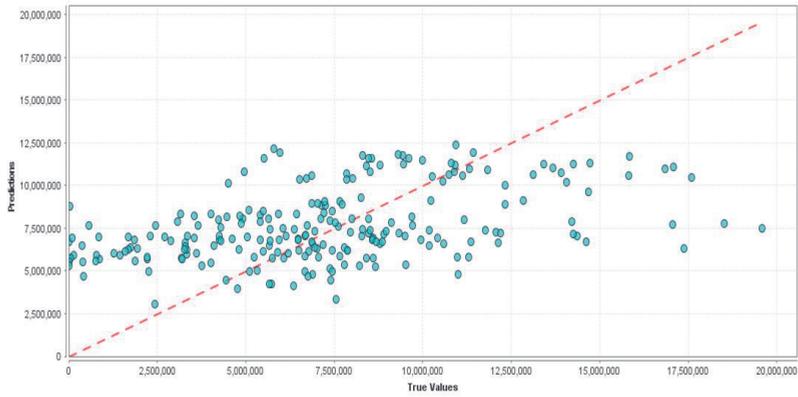


Figure 6. The scatter plot between production and prediction of production of Generalized Linear Model.

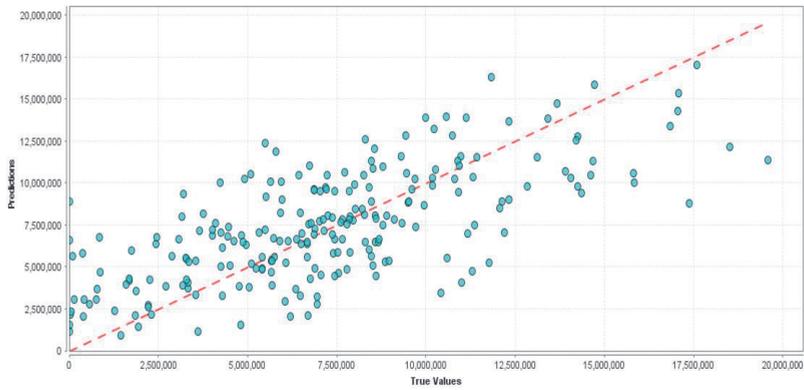


Figure 7. The scatter plot between production and prediction of production of Gradient Boosted Trees Model.

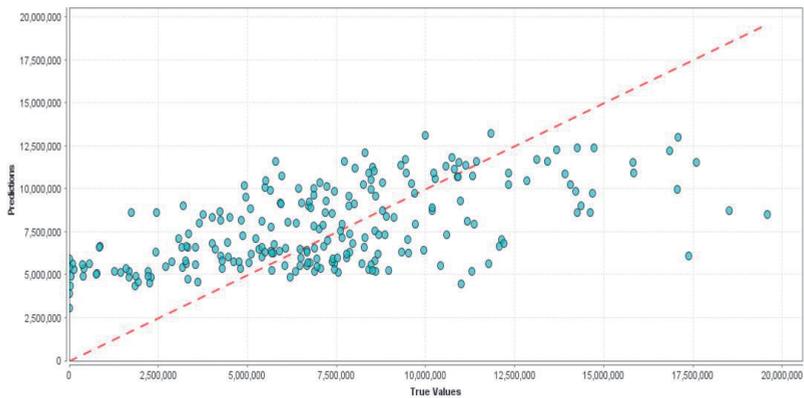


Figure 8. The scatter plot between production and prediction of production of Random Forest Model.

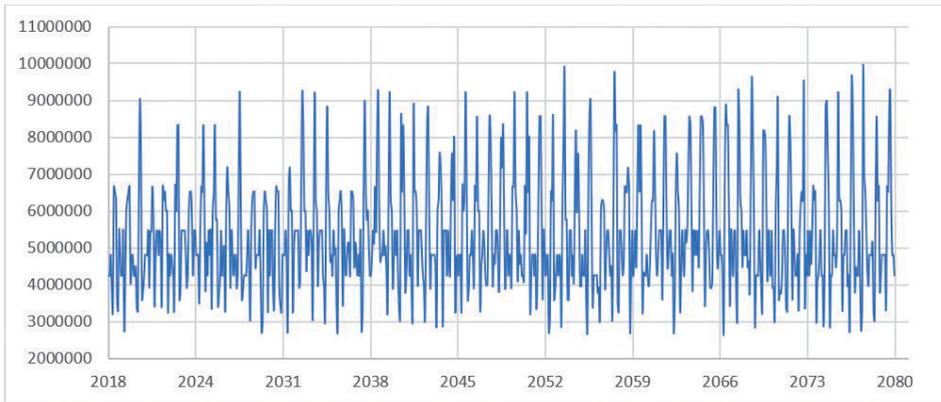
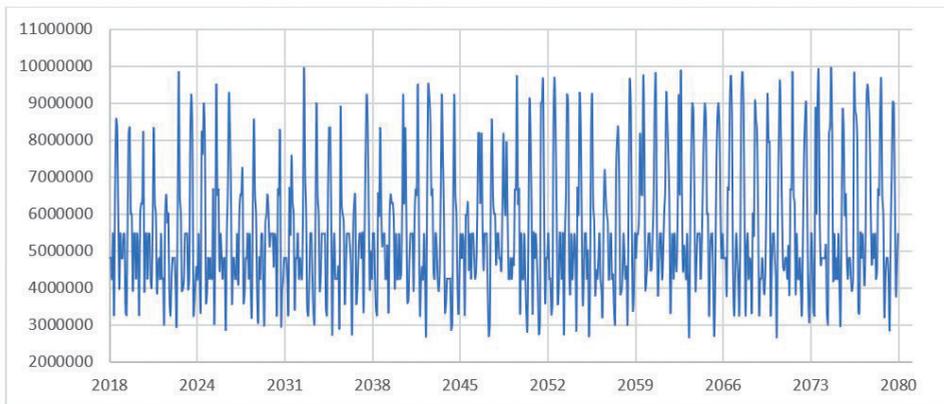
(a) GFDL 4.5*(b)* GFDL 8.5

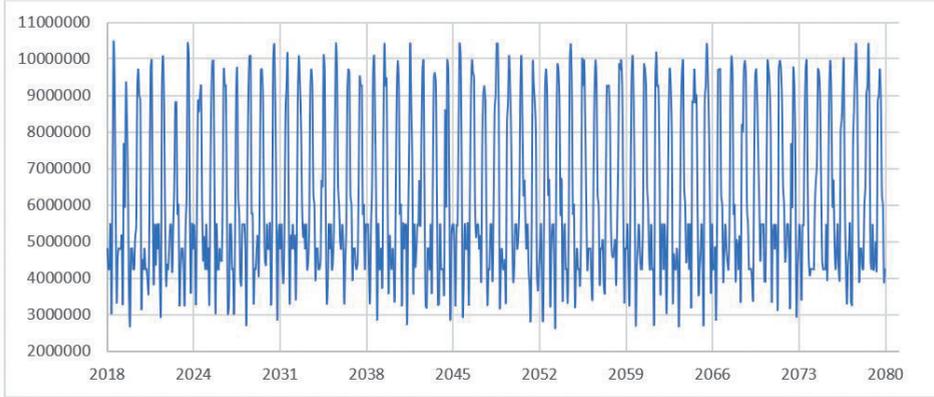
Figure 9. *(a)* and *(b)* The prediction productions of electricity (kWh) results from GFDL of precipitation and temperature data.

duction, more than GFDL 4.5 results during the same period. On the other hand, the results show a little difference between the two scenarios.

The prediction of the energy production for 4.5 and 8.5 scenarios using HadGEM model are shown in Fig. 10. The results of both scenarios are very close to each other with small difference during the time interval. The values range also are at the same level between 2,500,000 to 10,600,000 kWh.

The energy production in the GFDL 8.5 has increased faster than the GFDL 4.5. The range of values in the HadGEM 4.5 and HadGEM 8.5 are very close to each other. The trendlines are very close to the same range which indicates that

(a) HadGEM 4.5



(b) HadGEM 8.5

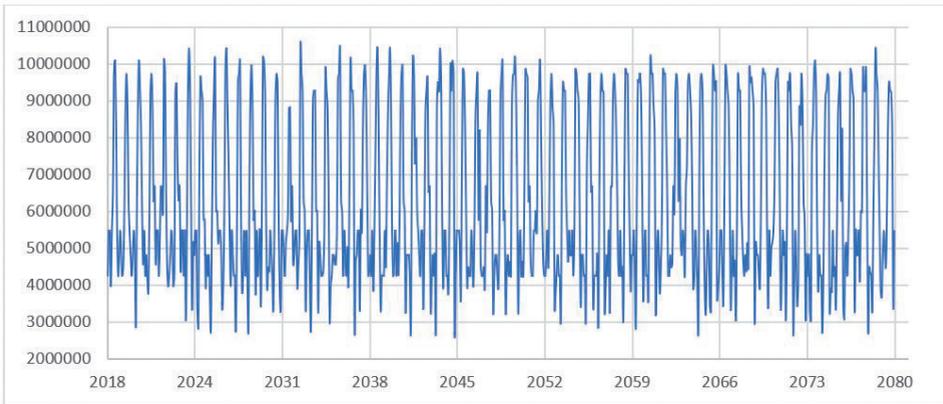


Figure 10. (a) and (b) The prediction productions of electricity (kWh) results from HadGEM of precipitation and temperature data.

the prediction of energy productions is very close based on the prediction of precipitation and temperature values. In Fig. 11, The prediction of energy production using MPI model shows that the values increase with time for the 8.5 scenario in comparison to the 4.5 scenario. In part b, the results show a high trend increase in comparison to the results of part a. This means that the trend of the 8.5 scenario is more than the trend of the 4.5 scenario. MPI 8.5 Scenario curve shows low values at the starting interval between 2018 to 2020. The high percentage of energy production values for all GCM 4.5 and 8.5 are in the same value level between 5,000,000 and 6,000,000 kWh. Although there are differences between precipitation and temperature data for GCMs, the prediction

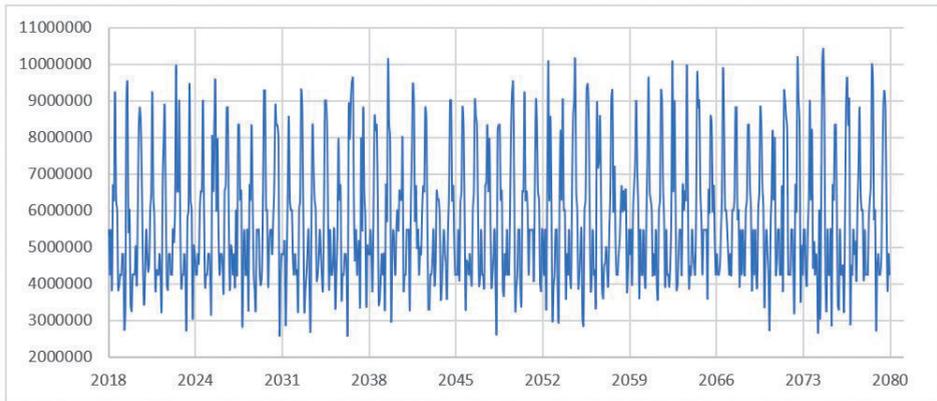
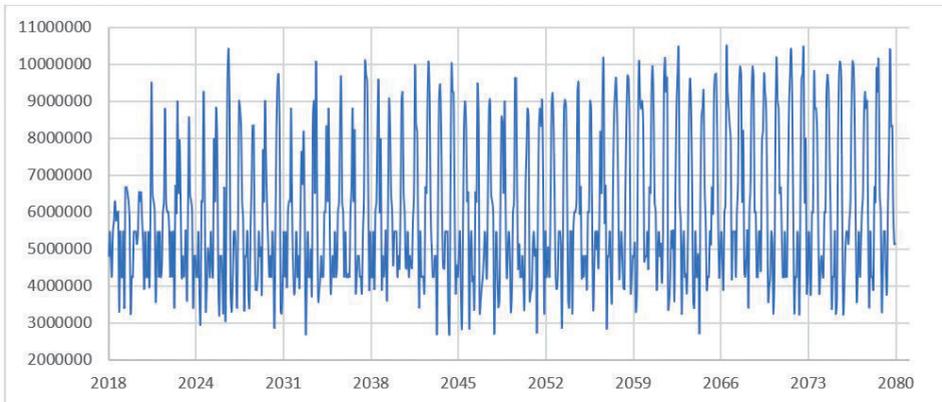
(a) MPI 4.5*(b)* MPI 8.5

Figure 11. *(a)* and *(b)* The prediction productions of electricity (kWh) results from MPI of precipitation and temperature data.

productions of electricity results are very near to each other. Each GCM represented two scenarios but the results were very near to each other for the interval between 2018 to 2020.

All the GCMs results show that the highest values were in the sudden raining season (July and August) and the lowest values were in the months of April and May. In the rest of the year, the curve is steady between the levels of 5,000,000 and 6,000,000 kWh. HadGEM results show a higher value than the other GCMs in the sudden raining season (July and August). Although all the GCMs showed an increased trend of energy production, this trend was different between the models.

5. Conclusions

Renewable energy mainly depends on local environmental conditions, such as precipitation, temperature, and rainfall-runoff ratios. Because of that, the expected power production heavily fluctuates, which makes estimating and calculating the feed-in into the power grid very challenging to be calculated and predicted. Hydropower is the primary renewable source for supplying clean energy, and its future contribution is anticipated to increase significantly.

According to the precipitation and temperature values of the GCMs, the paper presented the results of deploying Machine Learning Techniques in forecasting the amount of energy, which will be produced by Almus Dam and Hydroelectric Power Plant in Tokat, Turkey. Firstly, five models, Decision Tree (DT), Deep Learning (DL), Generalized Linear (GL), Gradient boosted trees (GBT) and Random forest (RF), were used to forecast the energy production of Almus HEPP using monthly hydroelectric power generation data from 1971 to 2013. The paper showed the differences between the five models and the quality of results in each model. DT, DL, GL, GBT and RF are utilized to forecast the hydropower production of the HEPPs. Five hundred and sixteen monthly data samples were used in each HEPP as data range from these HEPPs (Production, Total incoming water, and Incoming flow) to operate these models. The Precipitation and temperature data of the GCMs (HadGEM, GFDL and MPI) are used to forecast the energy production of the HEPPs for two scenarios (RCP 4.5 and RCP 8.5).

The forecast is calculated using only one machine learning model after checking the most accurate model between them (DT, DL, GL, GBT and RF). The checking was applied base on the absolute error and the correlation values. The highest cooperation and the lowest absolute error values for one of the machine learning models. The correlation values are 0.543, 0.401, 0.447, 0.717 and 0.637 for DT, DL, GL, GBT, and RF, respectively. The correlation value verified that Gradient boosted trees model gives more accurate results to the case study (Almus hydropower data). The Gradient boosted trees model was used for predicting the production of electricity. This study makes a clear indication about the hydropower production with time using three GCMs and two RCP scenarios in forecasting to understand the behavior of the system. The results show that there are small differences between the models which means that the predictions are going in similar directions at all these models.

References

- Awad, M. and Khanna, R. (2015): Machine learning, in: *Efficient learning machines*. Apress, Berkeley, CA, https://doi.org/10.1007/978-1-4302-5990-9_1.
- Beyene, T., Lettenmaier, D. P. and Kabat, P. (2010): Hydrological impacts of climate change on the Nile River basin: Implications of the 2007 IPCC climate scenarios, *Climate Change*, **100**, 3–4, <https://doi.org/10.1007/s10584-009-9693-0>.

- Breiman, L. (2001): Random forests, *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Breiman, L. (1997): *Arcing the edge*, Technical Report 486. Statistics Department, University of California, Berkeley.
- Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V. (2013): Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues, *Nat. Hazards Earth Syst. Sci.*, **13**, 2815–2831, <https://doi.org/10.5194/nhess-13-2815-2013>.
- General Directorate of State Hydraulic Works (2018): Almus Dam details, last available at <http://web.archive.org/web/20180225190623/http://www2.dsi.gov.tr/baraj/detay.cfm?BarajID=21>.
- Forrest, K., Tarroja, B., Chiang, F., Aghakouchak, A. and Samuelsen, S. (2018): Assessing climate change impacts on California hydropower generation and ancillary services provision, *Climatic Change*, **151**, 395–412, <https://doi.org/10.1007/s10584-018-2329-5>.
- Friedman, J. H. (1999): Greedy function approximation: A gradient boosting machine, p. 34, last available at http://biostat.jhsph.edu/~mmccall/articles/friedman_1999.pdf.
- Gökgöz, F. and Filiz, F. (2018): Deep learning for renewable power forecasting: An approach using LSTM neural networks, *Int. J. Energ. Power Eng.*, **12**, 412–416, <https://doi.org/10.5281/zenodo.1317172>.
- Huang, H. and Yan, Z. (2009): Present situation and future prospect of hydropower in China, *Renew. Sust. Energ. Rev.*, **13**, 1652–1656, <https://doi.org/10.1016/j.rser.2008.08.013>.
- Kankal, M., Akpınar, A., Komurcu, M. I. and Ozsahin, T. S. (2011): Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables, *Appl. Energ.*, **88**, 1927–1939, <https://doi.org/10.1016/j.apenergy.2010.12.005>.
- Kaygusuz, K. (2002): Sustainable development of hydroelectric power, *Energ. Source.*, **24**, 803–815, <https://doi.org/10.1080/00908310290086725>.
- Krušić, J., Marjanović, M., Petrović, M. S., Abolmasov, B., Andrejev, K. and Miladinović, A. (2017): Comparison of expert, deterministic and Machine Learning approach for landslide susceptibility assessment in Ljubovija Municipality, Serbia, *Geofizika*, **34**, 251–273, <https://doi.org/10.15233/gfz.2017.34.15>.
- Liaw, A. and Wiener, M. (2002): Classification and regression by random forest, *R. News*, **2**, 18–22.
- Melikoğlu, M. (2013a): Vision 2023: Assessing the feasibility of electricity and biogas production from municipal solid waste in Turkey, *Renew. Sust. Energ. Rev.*, **19**, 52–63, <https://doi.org/10.1016/j.rser.2012.11.017>.
- Melikoğlu, M. (2013b): Vision 2023: Forecasting Turkey's natural gas demand between 2013 and 2030, *Renew. Sust. Energ. Rev.*, **22**, 393–400, <https://doi.org/10.1016/j.rser.2013.01.048>.
- Mosavi, A., Salimi, M., Ardabili, S. F., Rabczuk, T., Shamshirband, Sh. and Várkonyi-Kóczy, A. R. (2019): State of the art of machine learning models in energy systems, a systematic review, *Energies*, **12**, 1–42, <https://doi.org/10.3390/en12071301>.
- Nelder, J. A. and Wedderburn, R. W. M. (1972): Generalized linear models, *J. R. Stat. Soc. A - G.*, **135**, 370–384, <https://doi.org/10.2307/2344614>.
- Sattari, M. T., Anli, A. S., Apaydin, H. and Kodali, S. (2012): Decision trees to determine the possible drought periods in Ankara, *Atmosfera*, **25**, 65–83.
- Smith, G., Mander, J. and Callenbach, E. (2012): *Nuclear roulette: The truth about the most dangerous energy source on Earth*. Chelsea Green Publishing Company, USA, p. 157, ISBN: 978160358435.
- TEDAS (Turkish Electricity Distribution Corporation) (2016): <http://www.tedas.gov.tr>.
- Yüksel, Ö., Kömürcü, M. I., Yüksel, I. and Kaygusuz, K. (2006): The role of hydropower in meeting Turkey's electric energy demand, *Energ. Policy*, **34**, 3093–3103, <https://doi.org/10.1016/j.enpol.2005.06.005>.
- Yüksel, I. (2008): Hydropower in Turkey for a clean and sustainable energy future, *Renew. Sust. Energ. Rev.*, **12**, 1622–1640, <https://doi.org/10.1016/j.rser.2007.01.024>.

SAŽETAK

Prognoza proizvodnje hidroenergije općim cirkulacijskim modelima primjenom tehnika strojnog i dubokog učenja (brana Almus, Turska)*Hesham Al Rayess i Asli Ülke Keskin*

Obnovljiva energija jedan je od najvažnijih čimbenika za razvijena i održiva društva. Međutim, njezina upotreba u elektroenergetskim sustavima može biti vrlo izazovna s obzirom na nepredvidljivost proizvodnje. Obnovljiva energija uglavnom ovisi o uvjetima okoline poput količine oborine, intenziteta otjecanja i temperature zraka. Zbog toga očekivana proizvodnja električne energije jako fluktuiraju, što prognozu i proračun njenog unosa u elektroenergetsku mrežu čini vrlo izazovnim zadatkom. Točno predviđanje proizvodnje energije iznimno je važno za proces upravljanja energijom. U ovom radu se predstavljaju rezultati primjene tehnika strojnog učenja u kratkoročnom predviđanju količine proizvedene energije na temelju rezultata općih modela cirkulacije (GCM) za branu i hidroelektranu Almus blizu naselja Tokat u Turskoj. Studija prikazuje upotrebu tehnika modeliranja u procesu prognoze proizvodnje hidroenergije pomoću prognoziranih mjesečnih podataka GCM-a o proizvodnji hidroelektrana u razdoblju od 2018. do 2080. Za prognozu proizvodnje hidroenergije korišteni su modeli: dijagrama odlučivanja, dubinskog učenja, generalizirani linearni, dijagrama pojačanih nagiba i dijagrama slučajnih grana. Vrijednost korelacije s modelom dijagrama pojačanih nagiba iznosi 0,717, što znači da je to najuspješniji model za korištene podatke. Model dijagrama pojačanih nagiba korišten je u svakom GCM-u za dva scenarija: RCP4.5 i RCP8.5. Rezultati pokazuju da postoje male razlike između modela, što znači da predviđanja idu u sličnim smjerovima za sve ove modele.

Ključne riječi: obnovljiva energija, vodna snaga, tehnike strojnog učenja, duboko učenje, opći cirkulacijski model, Turska

Corresponding author's address: Hesham Al Rayess, Ondokuz Mayıs University, Engineering Faculty, Civil Engineering Department, Samsun, Turkey; tel: +90 505 6928 238; e-mail: hesham.majed@hotmail.com (ORCID: 0000-0001-6983-145X).



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.