Climate Change and its Effect on the Energy Production from Renewable Sources – A Case Study in Mediterranean Region

Eralda Gjika[1*], Lule Basha[1], Arnisa Sokoli[1]

[1] Department of Applied Mathematics, Faculty of Natural Science, University of Tirana, Sheshi Nënë Tereza 4, Tirana 1010, Albania

* Corresponding author's e-mail: eralda.dhamo@fshn.edu.al

ABSTRACT

In terms of climate forecasting, the Mediterranean region is among the most difficult. It is correlated with the five significant subtropical high pressure belts of the oceans and is symbolized by dry and hot summer and cold and rainy winter. Due to its location in the area, Albania is particularly susceptible to climatic changes. It has been noted that summertime sees the greatest temperature increases. More intense heat waves that stay longer and occur more frequently are anticipated in the eastern Mediterranean. The seasonal patterns of precipitation have not changed, but the amount of rain has become more intense. The effects of climate change have drawn attention to various renewable energy sources, including solar and wind power. In this study, the changes and prospective in average temperature, rainfall, humidity, CO_2 emission and their impact in energy production were investigated. Several different models such as Auto Regressive Integrated Moving Average method; Prophet algorithm; ElasticNet Regularized Generalized Linear Model; Random Forest Regression models; Prophet Boost algorithm; have been built for the study and prediction of each variable. The appropriate models are used to determine the anticipated values of the indicators for a period of four years. The prediction shows an increase in CO_2 emission which leads to a decrease in energy production by hydropower. These findings suggest the use of other renewable sources for energy production in the country and the Mediterranean region.

Keywords: energy production, temperature, rainfall, CO_2 emissions, machine learning models.

INTRODUCTION

Climate changes have had an impact not only on life processes but also on different areas of the economy. Research studies have shown that the world's surface air temperature increased an average of 0.6° Celsius during the last century. This process has been accelerated by the burning of fossil fuels which releases carbon dioxide and other greenhouse gases into the atmosphere associated with the development of industry and population growth. The climatic factors such as: average temperature, precipitation, and CO_2 emission, are linked with economic and financial developments in different climatic regions. Most significant hazards associated with climate change are: extreme temperature; windstorms; drought; and floods. Energy is one of the most vulnerable

sectors. Thus, these hazards are important and directly linked to the production of electricity by renewable source and especially hydropower plants. Since the main source for energy production by HPP involves water inflow from rivers, snow melt and rainfalls, the average temperature and also CO_2 emissions may directly affect the amount produced by hydropower. Hydropower generation is one of the most favourable sources of energy since it has many advantages, such as the low cost and the absence of greenhouse gas generation.

The seasonal nature of some of the climacteric factors that affect the amount of energy produced by HPP requires that the predictive models are able to capture the visible but also the hidden seasonality. Short-term and long-term prediction scenarios of climate change and its effect have been studied by many researchers through the last decade.

This study highlighted the challenges and opportunities concerning the impact of climacteric factors on the amount of energy produced by hydropower. The results are based on the data for a Mediterranean country, such as Albania, but the models may be useful to other countries in the region. The nation is prone to a wide range of environmental dangers, including weather-related risks, geological risks like earthquakes, and other catastrophic phenomena. Severe storms, floods, heat waves, and wildfires become increasingly frequent and unexpected. The occurrence of "heat waves" has increased recently; 74% of all incidents have taken place since 1996. Because the nation's river and stream systems are important suppliers of water during rainstorms, it is extremely vulnerable to floods.

Since the 1970s, the average annual temperature in Albania has risen by 1.5°C. However, very limited information on climate effects on power generation, transmission and distribution is currently available for the region and especially Albania. Related to the monthly average temperature summer and autumn (from May to October) show stable average temperatures over the years, while the spring and winter months have more frequent extreme temperatures in their history, see Figure 1. In recent years, a significant increase in the average temperature for the winter months has been observed. The same behaviour is noted for the average rainfall.

The summer months are the ones with the least amount of rain in the history of summer observations, see Figure 2. In turn, the autumn and winter months are represented by the highest amount of precipitations. Over the years, it has been observed that the amount of rainfall for the autumn season has decreased. Extreme values are present especially for the first and last months of the year. Seasonal patterns are also clear in both temperature and rainfall.

The area has developed hydropower facilities as a source of electricity because of the wet winters and a vast river network in the region. Albania is one of the nations whose primary source of electricity comes from hydropower plants (almost 80% of country production). A severe drought that lasted for 30 years drastically decreased the amount of water in the river cascade. The high amount of rainfall during the wet months also increases the amount of inflows in the cascade of rivers which are the primary source of energy production. In this way, the amount of energy produced during this season increases significantly and meets the needs of the local market and during the last years also for the region.

The correlation between temperature and rainfall is significant especially for autumn (cor=-0.457), spring (cor=-0.372) and summer (cor=-0.364). For winter a positive value close to zero is observed (cor=0.155), Figure 3.

Many classical prediction models have been used in the literature of climate change data more advanced models, such as machine learning models, are also needed. Franco et al., 2008 have estimated the impact of regional climate change on
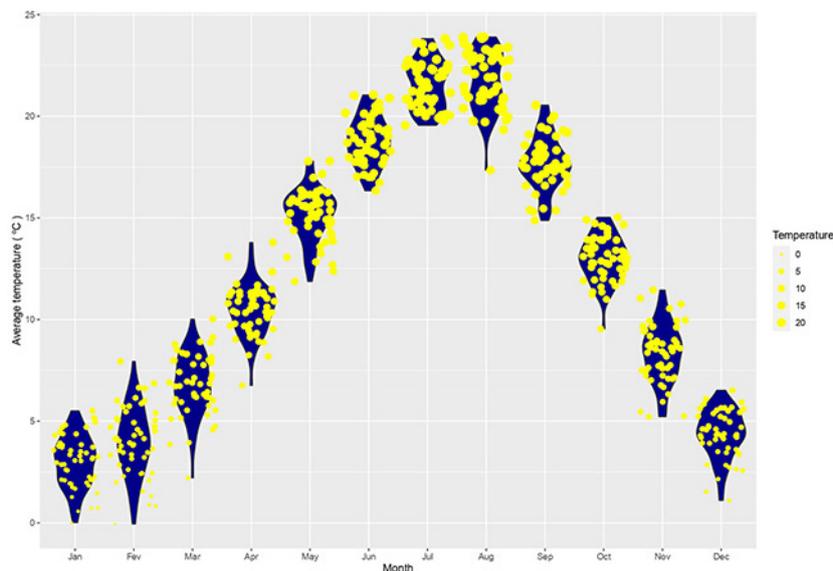

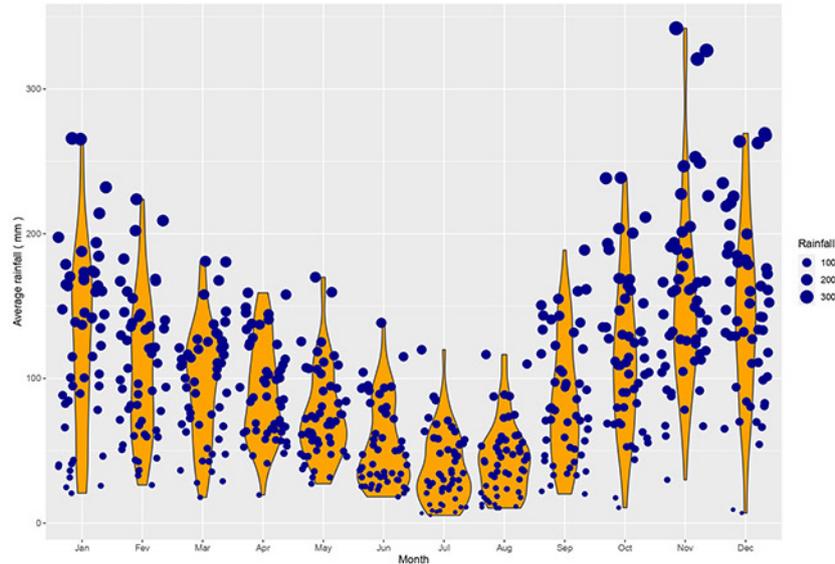**Figure 1.** Temperature distribution by month, in Albania

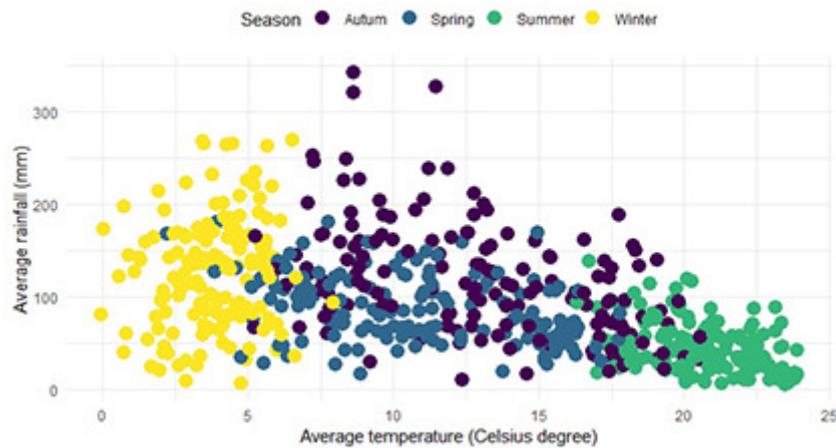**Figure 2.** Rainfall distribution by month, in Albania

Figure 3. Correlation between temperature and rainfall

electricity demand in California using historical relationships between peak demand and temperature. Another view of short- and long-term impacts of climate change on electricity demand is presented by Emodi et al. (2018). They used an autoregressive distributed lag (ARDL) model with monthly data from Australia. The results of the study showed that electricity consumption will indicate a slight increase in winter as a result of the significant decrease in temperatures in this period and the peak consumption will be in the summer season, when the temperatures are higher. Wadsack and Acker (2019) conducted production cost modelling to examine climate change and future power systems. They studied the importance of energy storage in reduced-hydropower systems in the American Southwest.

Mason et al. (2018) have focused their attention on machine learning techniques, modified neural networks in particular to obtain short term prediction of power demand, wind power generation and carbon dioxide intensity levels in Ireland. They showed that machine learning models provide fast convergence, more accurate predictions and robust performance.

Zhang et al. (2019) attempted to estimate the effect of temperature fluctuations on electricity consumption in rural area in China due to weather changes, such as annual average temperature and annual average precipitations. They used an econometric model to investigate the relation of climate change and energy demand. Taking into account the volatility of climacteric time series and energy produced by hydropower it has

considerably increased the need of using hybrid models to improve the quality of predictions. Gjika et al. (2019) presented ensemble models to obtain energy prediction models: Hybrid models based on LSSVM method; improved Hybrid models with SARIMA and ETS forecasting model. The LSSVM hybrid model showed more accurate prediction. Hale and Long (2021) used time series models, ARIMA and exponential smoothing, to develop a prediction of Missouri's annual electricity generation. The study showed that ARIMA exhibited superior performance. The model presented provides a univariate time series prediction of annual electricity generation using publicly available data. In their study, Guo et al. (2021) have studied the effects of climate change on hydroelectric power generation. They have used artificial neural network (ANN) optimized by Improved Electromagnetic Field Optimization (IEFO) algorithms to obtain prediction of energy demand.

Eysenbach et al. (2021) analyzed important factors for generation and load of Texas' power grid. They constructed different statistical models and machine learning with the data which included: weather, population, location, and time in its models and forecasts, all of which are autoregressive components of historical load data. Gjika et al. (2021) analyzed the seasonal pattern of climacteric factors: precipitation, average temperature, and water inflow affecting energy production. They considered different statistical learning methods for energy prediction: ARIMA, ETS, NN, TBATS, STLM, etc. and comparing the performance of the models they agreed for the data that neural networks provide good forecasts for monthly energy produced by hydropower. Gjika and Basha (2022), classical time series and deep learning models for energy load prediction. They evaluated the impact of climacteric factors on energy load using hourly and daily time series for a period of three years. On the basis of the frequency of the data, the models considered showed competitive performance for short-and medium-term energy load forecasts. Lahouar and Ben Hadj Slama (2015), proposed a load prediction model of one day ahead with resolution of one hour, using regression random forests. The main variables used are: season, temperature, type of the day and hourly load. The models proposed offered accurate and effective long term prediction. A forecasting model based on the random forest

algorithm was created by Meng and Song (2020) for three categories of winter days in North China. The daily power generation at the Zhonghe PV station, which is situated in the middle of North China, is forecasted using the proposed model as well as three additional approaches in order to assess its performance. Serras et al. (2019), combined random forests with physics-based models to forecast the electricity output of the Mutriku wave farm on the Bay of Biscay.

In this article a case study employing average temperature, rainfall, CO_2 emissions, and their impact on energy production from renewable sources is described.

METHODOLOGY

The main solution to all issues with predictive modelling is frequently said to be machine learning and deep learning techniques. Different machine learning techniques, as well as traditional time series forecasting techniques have been used in this study. The data have the form of $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_k \in \mathbb{R}$ is the response variable and $x_k = (x_{k,1}, \ldots, x_{k,p})^T \in \mathbb{R}^p$ is the vector of independent variables.

Modelling

Auto regressive integrated moving average

By modelling the correlations in the data, the ARIMA methodology is a statistical technique for assessing and creating a forecasting model that accurately depicts a time series, Box and Jenkins, 1976. In order to generalize the forecast and boost prediction accuracy while maintaining the model's parsimony, ARIMA models only require the past data of a time series. The full model can be written as:

$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \\ + \theta_1 \varpi_{t-1} + \cdots + \theta_q \varpi_{t-q} + \varpi_t \qquad (1)$$

where: p – the order of the autoregressive part,
d – the degree of first differencing involved, q – the order of the moving average part; ϖ_t – is white noise.

Past data are typically used as predictor variables in time series modelling. However, other factors that have an impact on the target variables are occasionally required. Regression modelling

is necessary to include those variables. To capture sophisticated patterns, dynamic regression will be employed instead; in contrast to conventional regression models, this one uses ARIMA to describe residuals. The regression model error and the ARIMA model error are the two error factors in this model. The faults in the ARIMA model are considered to be white noise (Harris and Sollis, 2003).

Prophet algorithm

Prophet is a well-known local Bayesian structural model used for time series prediction. It is a technique used especially when dealing with time series data based on an additive model where non-linear trends are coupled with yearly, weekly, and daily seasonality, as well as holiday effects. Strongly seasonal time series and multiple seasons of historical data are ideal for it (Taylor and Letham, 2018). Since the main concern in these time series are outliers, Prophet typically manages outliers well and is robust to missing data and changes in the trend.

The process is fundamentally an additive regression model with four major parts, of the form

$$y_t = l(t) + p(t) + h(t) + \varpi_t \tag{2}$$

where: $l(t)$ – denotes a logistic growth curve or piecewise-linear trend.

It uses change points from the data to automatically identify trends that have changed. The seasonal component of $p(t)$, which describes the distinct seasonal patterns, is made up of Fourier terms for the pertinent periods. Holiday impacts are inserted as straightforward dummy variables and are described by $h(t)$ and white noise is present by ϖ_t.

Prophet boost algorithm

A well-liked and effective version of the gradient boosted trees technique is XGBoost. Regression trees serve as the weak learners when utilizing gradient boosting for regression, and each one of them associates each input data point with a leaf that holds a continuous score. A convex loss function and a penalty term for model complexity are combined to form a regularized objective function that is minimized using XGBoost. Because the loss when introducing new models is minimized, the technique is known as gradient boosting. In order to achieve the best of both worlds, machine learning and prophet automation, the Prophet Boost algorithm combines XGBoost with Prophet. Prophet modelling of the univariate series is the initial step in the algorithm's operation. Following that, the Prophet Residuals are regressed using the XGBoost model while employing regressors provided by the pre-processing recipe. XGBoost is a popular and efficient open-source implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models.

Elastic-net regularized generalized linear model

One of the most popular methods for inferential modelling is the use of generalized linear models (GLMs). They are particularly useful tools for expressing causal inference to stakeholders since they are straightforward and simple to understand. A popular regularization technique called elastic net regularization removes irrelevant and highly correlated information, which can be detrimental to accuracy and inference. These two techniques are a valuable addition to any data scientist's toolset. To efficiently compute a model's coefficients for any link function, not only the simple ones, Tay et al. (2018), published a research that makes use of cyclic coordinate descent.

The response can be modelled as a linear combination of the covariates, according to the ordinary least squares model. The residual sum of squares is minimized in order to estimate the parameters. In the last two decades, regularization techniques have been the subject of extensive research. Of particular interest are the elastic net that minimizes the total of the residual sum of squares and a regularization term that combines ℓ_1 and ℓ_2 penalties:

$$(\hat{\beta}_0, \hat{\beta}) = \underset{(\beta_0, \beta) \in \square^{p+1}}{\arg\min} \left[\frac{1}{2n} \sum_{k=1}^{n} (y_i - \beta_0 - x_k^T \beta)^2 + \right.$$

$$\left. + \gamma \left(\frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \right) \right] \tag{3}$$

where: γ and α – the tuning parameter and a higher level hyper parameter, respectively.

Three components make up a generalized linear model: a linear predictor, a link function, and a variance function that depends on the mean. The residual sum of squares term in Equation 3 is changed to a negative log likelihood term in order to apply the elastic net to GLMs:

$$(\hat{\beta}_0, \hat{\beta}) = \arg\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} \left[-\frac{1}{n} \sum_{k=1}^{n} \ell(y_k, \beta_0 + x_k^T \beta)^2 + \right.$$

$$\left. + \gamma \left(\frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \right) \right] \quad (4)$$

where: $\ell(y_k, \beta_0 + x_k^T \beta)^2$ – the log-likelihood term connected to observation k.

The approach used to minimize this objective function is the same as that used for GLMs. The main distinction is that a penalized weighted least squares problem is addressed in place of a weighted least squares problem in each iteration (Friedman et al., 2010).

Random forest regression models

The most well-known classification algorithm, Random Forest (Breiman, 2001), can perform both classification and regression. Large amounts of data can be accurately classified using it. The fact that the method combines decision trees gives it the name "Random Forest." Each tree in the "forest" is reliant on the values of a random vector sampled independently from the others using the same distribution. Each one has been given the best chance of growth. The Random Forest principle states that although though individual trees may be "poor learners," when they are all combined, they might make up a single "strong learner." Random selection of input variables enables the achievement of this objective in the tree growth process. After T such trees $\{K(x; \Theta_t)\}_1^T$ are grown, the random forest (regression) predictor is:

$$\hat{f}_{rf}^T(x) = \frac{1}{T} \sum_{t=1}^{T} K(x; \Theta_t) \quad (5)$$

where: Θ_t – characterizes the *t-th* random forest tree in terms of split variables, cut points at each node, and terminal-node values.

Evaluation metrics

The performances of the models presented in this study were evaluated in terms of a number of metrics. The selection of the most accurate model is made by analysing and comparing error measurements and information criteria for proposed models, as well as extending personal judgment to the advantages offered by each model based in the nature of the data. The metrics used to assess and compare the various methods are:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{X}_t - X_t| \quad \text{maximum absolute error} \quad (6)$$

$$ME = \frac{1}{n} \sum_{i=1}^{n} (\hat{X}_t - X_t) \quad \text{mean error} \quad (7)$$

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{X}_t - X_t|}{|X_t|} \right) \cdot 100\% \quad \text{mean absolute percentage error} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{X}_t - X_t)^2} \quad \text{root mean square error} \quad (9)$$

$$MASE = \frac{MAE}{MAE_{in-sample, naive}} \quad \text{mean absolute scaled error} \quad (10)$$

Many arguments and discussions of using the appropriate accuracy measurements of the model are presented in the literature. On the basis of these discussions and the nature and complexity of the data in reference of (Hyndman & Koehler, 2006) MASE offers a straightforward indication on the relative model performance compared with the naïve benchmark. It is a scale-independent measure where a value of less than one indicates that the performance of the model is better than the naïve benchmark on average, whereas a value greater than one indicates the opposite. What is important is the fact that this critical value should not conclude the performance of the model but further analysis is suggested.

RESULTS AND DISCUSSIONS

The objective of this paper was to study the changes in average temperature, rainfall, CO_2 emission and their impact in energy production. The data used in this work are obtained from official publications, The Word Bank and Databank. In the first step of this work, average monthly temperature and rainfall, annual CO_2 emissions in Albania were studied, for the period 1970–2021. Statistical analysis was performed for each of the indicators. Then, several different models have been built for the study and prediction. The authors have used classical models such as Auto Regressive Integrated Moving Average method, then machine learning algorithms: Prophet algorithm, local Bayesian structural time series model; elastic-net regularized generalized linear model; random forest regression models; prophet boost algorithm. The correctness of each model was also assessed using performance metrics. The appropriate models are used to determine the anticipated values of the indicators for a period of
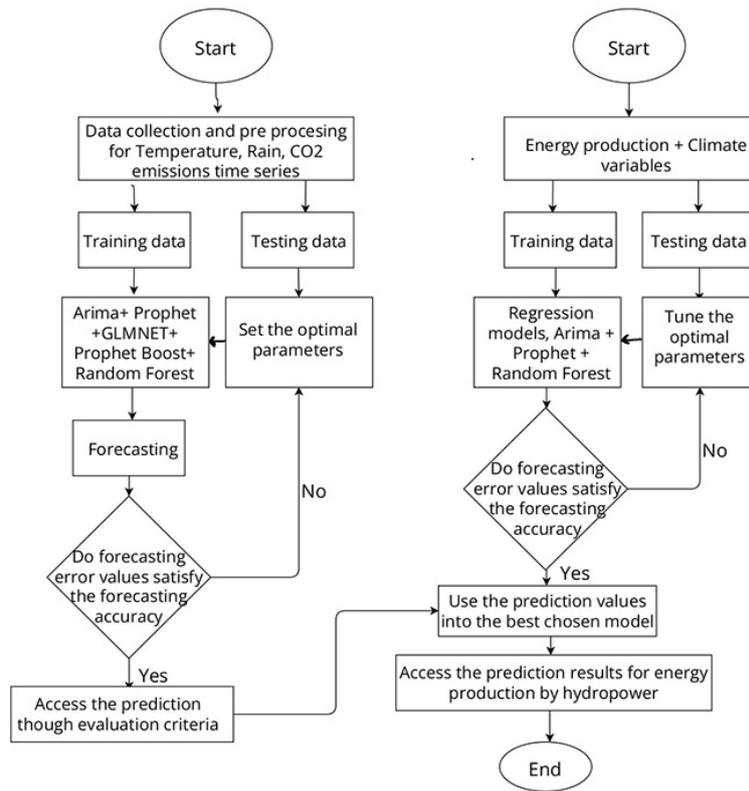
Figure 4. The flow diagram of the study

four years after choosing the optimal model for each explanatory variable.

In the following phase, the effects of temperature, precipitation and CO_2 emissions on the annual energy produced by hydropower for the years 1990 through 2021 were examined. The predicted values received for each of the variables are placed in the best chosen model, in order to obtain the annual forecast of energy production by hydropower, for 2022 to 2025, see Figure 4.

Temperature, rainfall and CO_2 emissions

The mountains, hills, and coastline of Albania make up the majority of its topography, and the nation's geology and climate contribute to an enormous network of rivers and lakes. Since the 1970s, the average annual temperature in Albania has risen by 1.5°C. Intensity, length, and frequency of heat waves are predicted to rise in the eastern Mediterranean, maybe up to six to eight times per year. Summer temperatures are comparable in all coastal locations of Albania; however, the northern half of the country's coastal zone normally experiences lower winter temperatures than the central and southern zones. Albania's average annual temperature from 1970 to 2021, is 12.43°C; the year with the lowest

annual temperature, 10.6°C, was in 1977, and the year with the highest, 13°C, was in 2018. It has been noted that the summer months of June, July, August, and September have the greatest temperature increases. The months of April and November are also becoming warmer. In the past 20 years, September and November have had greater temperature variability, while March has experienced substantially less.

First, the time series were explored examining its components and seasonality structure, and a correlation analysis was carried out to identify the main characteristics of the series. The data displays a clear rising trend. The plot displays a clear seasonal pattern, and the random component seems to be dispersed randomly. Then, the data were divided into two part the train (80%) and test (20%) data, Figure 5, which also present the time series of monthly temperature from 1970 to 2021.

Several different models have been built for the study and prediction of temperatures. After the models were constructed, each one was tested, comparing the models' accuracy to performance metrics and real vs. projected plots. An accuracy table is created for the models in order to evaluate and contrast their performance, see Table 1. From the results of accuracy table, it can be concluded that the Prophet algorithm is the best model for
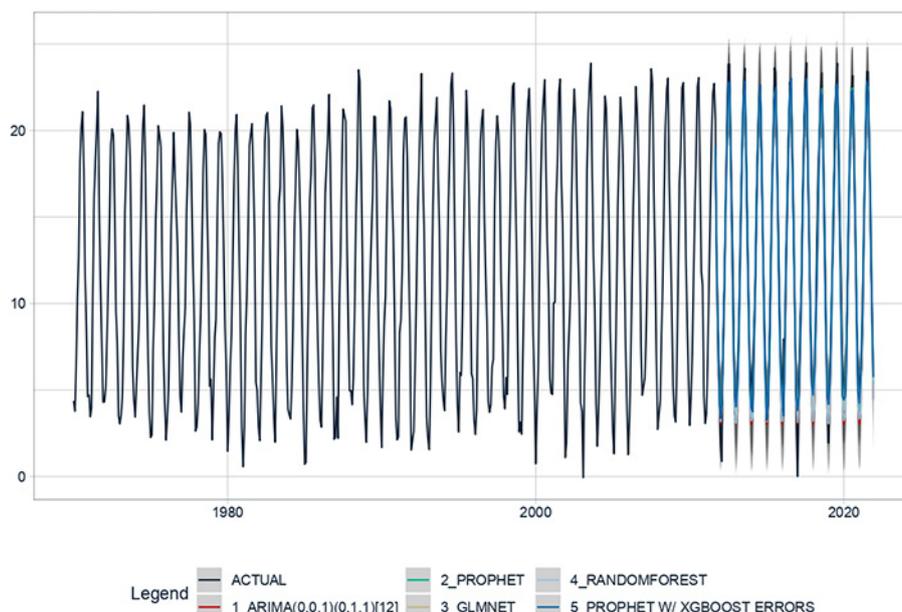
Figure 5. Monthly temperature in Albania, 1970–2021, and performance of models

Table 1. Accuracy table of models for temperature estimation

Model	MAE	MAPE	MASE	SMAPE	RMSE
1. ARIMA(0,0,1)(0,1,1)[12]	1.07	259.28	0.31	14.67	1.35 0.97
2. PROPHET	0.92	315.44	0.27	12.37	1.19 0.97
3. GLMNET	0.95	310.46	0.28	12.59	1.23 0.97
4. RANDOMFOREST	1.23	267.05	0.36	15.96	1.50 0.97
5. PROPHET W/ XGBOOST ERRORS	1.01	321.2	0.30	13.69	1.29 0.96

our data. On the basis of this model, Figure 6 give the predictions of temperature for 48 months for years 2022 and 2025, with mean annual temperature 13.18;13.22;13.28 and 13.31 Celsius degree, respectively.

Most of Albania's precipitation has fallen in its western regions, notably in the northwest. Although there is considerable intra-seasonal fluctuation in Albania's precipitation, since the 1970s there has been a very modest (but statistically insignificant) decline in mean annual precipitation. The number of rainy days each year has increased. Although seasonal precipitation patterns as a whole show little change, rainfall has become more intense. The amount of infrastructure that needs to be maintained and ready for use in managing flood waters are both impacted by the increased intensity of rainfall.

The regional variation of the measured average monthly rainfall for the period 1970–2021 is shown in Figure 7, together with models performance. There are the graphs of the actual vs forecasting values in the figure to follow, for each

model. The year 2011 recorded the least amount of rainfall, with an annual rainfall of 856.01 mm, while the year 2010 recorded the most, with an annual rainfall of 1470.31 mm. The average annual rainfall in Albania during the past 50 years has been 1136.03 mm.

The fundamentals of machine learning models vary greatly. Therefore, before selecting the optimal model, it is crucial to examine the predictive ability of at least two models. Thus, for the built models, for rainfall estimation and forecasting, an accuracy table was built, see Table 2. It was shown that for the given dataset, Prophet algorithm performs better than other models.

This model is used to make predictions of rainfall for 48 months for years 2022 and 2025, with annual rainfall 1189.7 mm, 1190.1 mm, 1190.7 mm, and 1191 mm, respectively, see Figure 8.

In Albania, the CO_2 emissions in 2020 decreased by 0.473 megatons, or 8.48%, from those in 2019. According to Figure 7, Albania's per-capita CO_2 emissions in 2020 were 1.73 tons, a drop over the previous year. Since 2010, both the
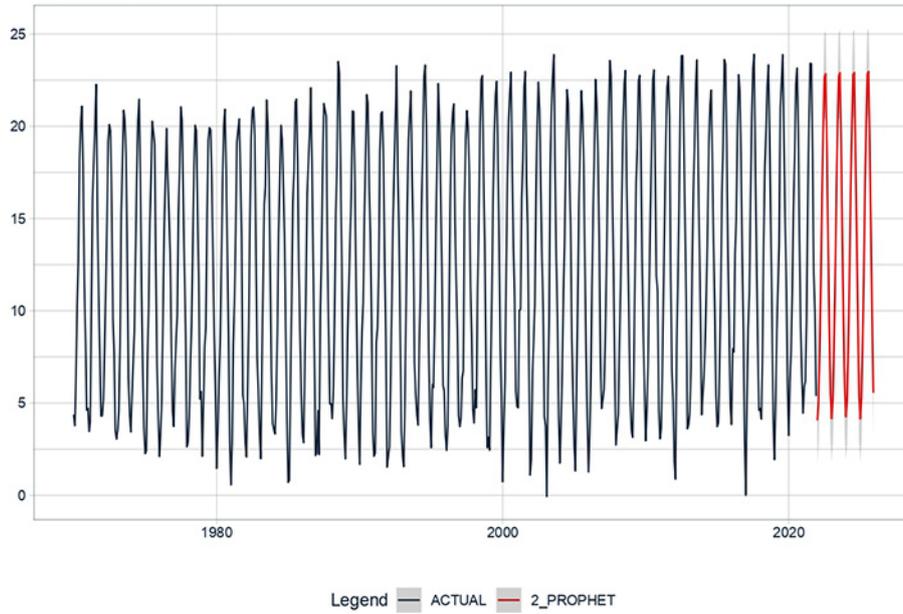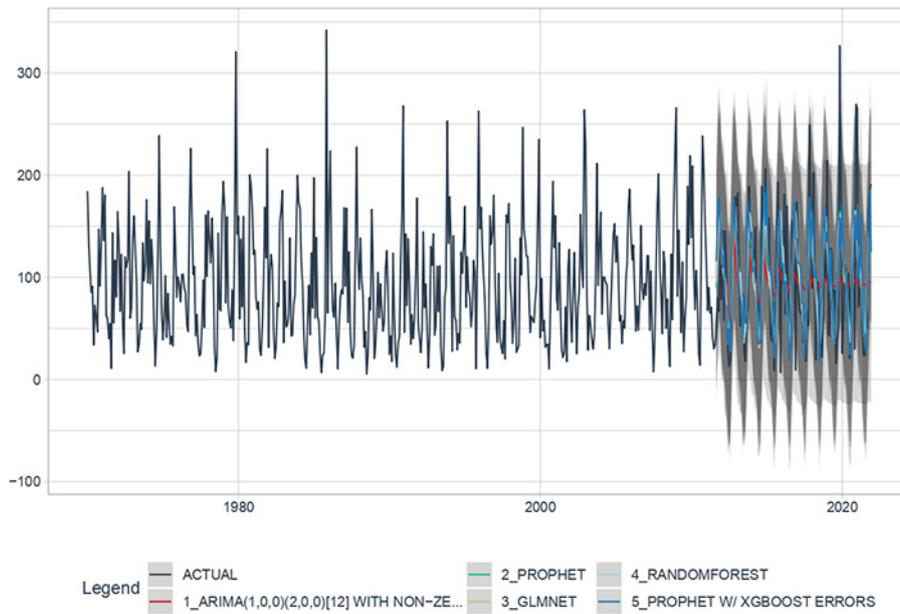
Figure 6. Temperature forecast

Figure 7. Monthly rainfall in Albania, 1970–2021, and models performance

Table 2. Accuracy table of models for rainfall estimation

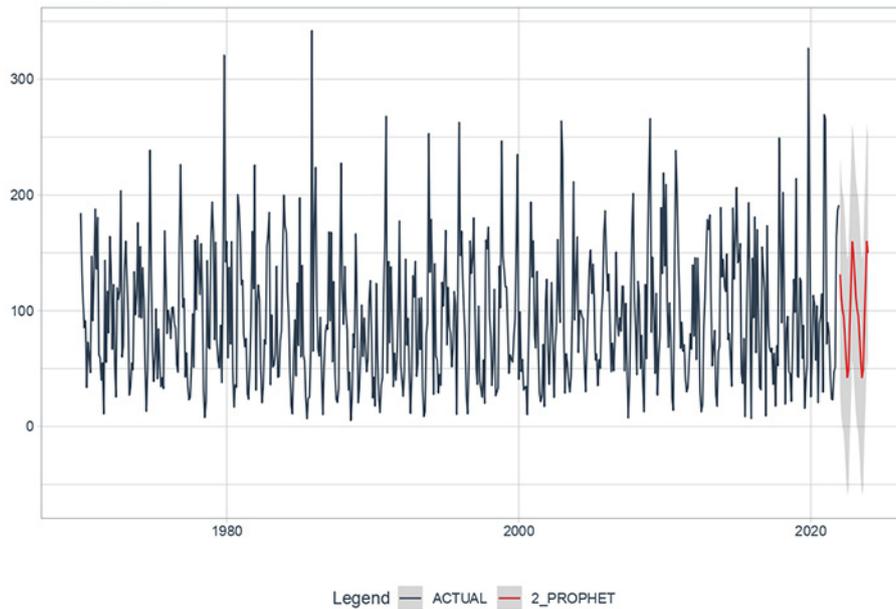| Model	MAE	MAPE	MASE	SMAPE	RMSE
1. ARIMA(1,0,0)(2,0,0)[12] WITH NON-ZERO MEAN	46.91	98.33	0.83	52.13	59.74
2. PROPHET	40.42	94.17	0.72	45.55	51.8
3. GLMNET	40.76	87.3	0.72	46.44	52.24
4. RANDOMFOREST	41.44	95.05	0.73	46.14	53.13
5. PROPHET W/ XGBOOST ERRORS	44.67	93.31	0.79	49.98	55.66

Figure 8. Rainfall forecast

evolution of CO_2 emissions and per capita emissions have increased. On the other hand, the CO_2 emissions per \$1000 of GDP have decreased. Both total and per capita carbon dioxide emissions have increased during the past five years, Figure 9.

For CO_2 emissions, the same models that were created for temperature and precipitation were also used, with training and testing sets of data. After the models were built, they were assessed, and the best model was selected. The forecasts were generated using each model and evaluated against the baseline.

According to the accuracy table, Table 3, the Prophet Boost algorithm is the best model for the considered data. This model has been used to make predictions of the CO_2 emissions for years 2022:2025, see Figure 10.

Machine learning random forest models to investigate the importance of temperature, rainfall and CO_2 emissions in energy production by hydropower

Albania is blessed with a wide range of energy resources, including coal, oil and gas, hydropower, biomass from natural forests, and other renewable energy sources. Albania is already susceptible to changes in precipitation, as seen by the severe energy shortages caused by the 2007 drought. In addition to harming coastal infrastructure, landslides and floods in the plains and lowlands in previous years also severely damaged infrastructure. The potential for hydropower in Albania is likely to decline as a result of rising temperatures, which are also anticipated to modify the seasonal need for heating, cooling, and refrigeration. As the nation develops, there will be a growing demand for energy from individuals, businesses, and the overall economy, which could exceed current energy production and transmission facilities. Albania has already committed to building at least three new solar photovoltaic power facilities starting in 2020 to accommodate this demand.

The next stage of the analysis looks at what impact temperature, precipitation, and CO_2 emissions have on the annual energy generated by hydropower from 1990 to 2021. The regression

Table 3. Accuracy table of the models for CO_2 emissions

Model	MAE	MAPE	MASE	SMAPE	RMSE
1. ARIMA(0,1,0)	0.41	8.7	0.89	8.39	0.45
2. PROPHET	1.76	36.62	3.83	45.02	1.79
3. GLMNET	2.29	47.6	4.96	64.85	2.41
4. RANDOMFOREST	1.32	27	2.86	31.52	1.37
5. PROPHET W/ XGBOOST ERRORS	0.34	6.75	0.74	7.15	0.46

Figure 9. Annual CO_2 emissions in Albania, 1970–2021

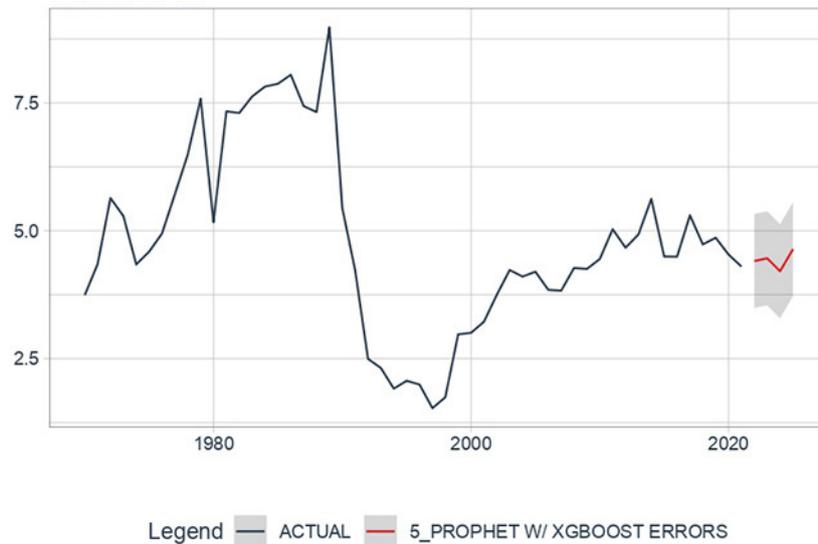Legend — ACTUAL — 5_PROPHET W/ XGBOOST ERRORS

Figure 10. CO_2 emissions forecast

models used for this purpose are regression model with ARIMA errors, Prophet algorithm and Random forest regression model. Table 1 gives the performance metrics for each of the models

From the results in Table 4, it can be seen that the model that best suits the data is the random forest model. To fine-tune the parameters, we created a unique random forest model. By comparing the results of several parameter combinations, it was possible to determine that 500 trees and a maximum of 80 nodes were the optimal parameters for the adopted model. The degree to which predictor factors reduced node impurity during split selection serves as a measure of their effectiveness.

Table 4. Performance metrics for each of the models

Model	Mape	Mase	rsq
Regression model with ARIMA errors	19.20	0.86	0.39
Prophet	18.45	0.83	0.53
Random forest	14.54	0.65	0.67

The variables' relative importances are depicted in the plot in Figure 11. The multi-way importance plot displays the relationship between three important metrics and identifies the top variables for each of these metrics, which are based on the forest's tree structure: the number of trees whose roots are divided on the variable; the mean depth of
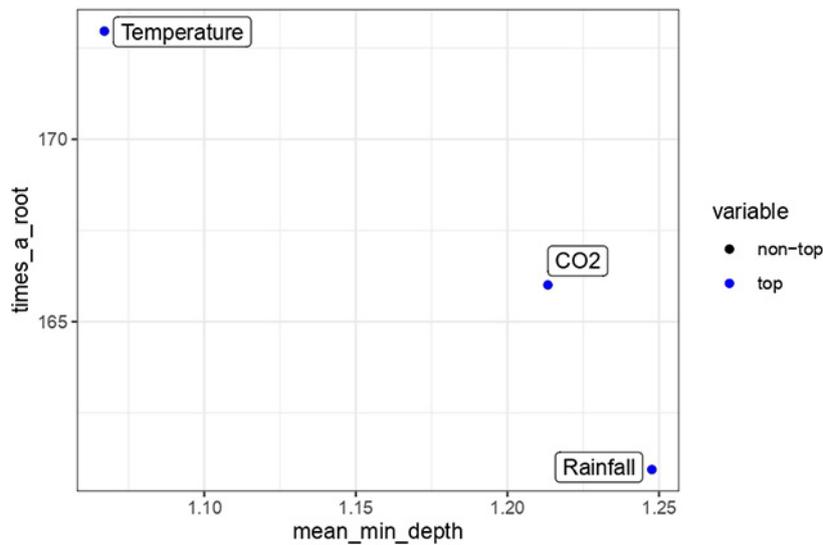
Figure 11. Multi-way importance plot

the first split on the variable; and the overall number of nodes in the forest that split on the variable. The behavior of energy generation by hydropower is clearly influenced by all of the variables, as is shown. The most important factor is Rainfall, followed by CO_2 emissions and temperature.

The distribution of minimal depth among the trees in your forest is depicted in Figure 12. The scale of the X axis ranges from zero to the maximum number of trees, in this case 500, in which any variable was used for splitting, and the mean of the distribution is indicated by a vertical bar with a value label on it. The depth of the node that splits on a variable and is the furthest from the

tree's root is the minimal depth for that variable in the tree. If it is low, this variable is used to partition a large number of observations into groups, in the presented study this variable is Temperature.

The random forest constructed model is used to anticipate the yearly hydropower energy output from 2022 to 2025 using the predicted values for temperature, rainfall, and CO_2 emissions. The results are plotted in Figure 13, with values 6239200; 6260434; 5452307 and 5870179 MWh, respectively. As it can be seen from the results, it is expected that there will be a decrease in the production of electricity from hydropower in Albania in the next 4 years.
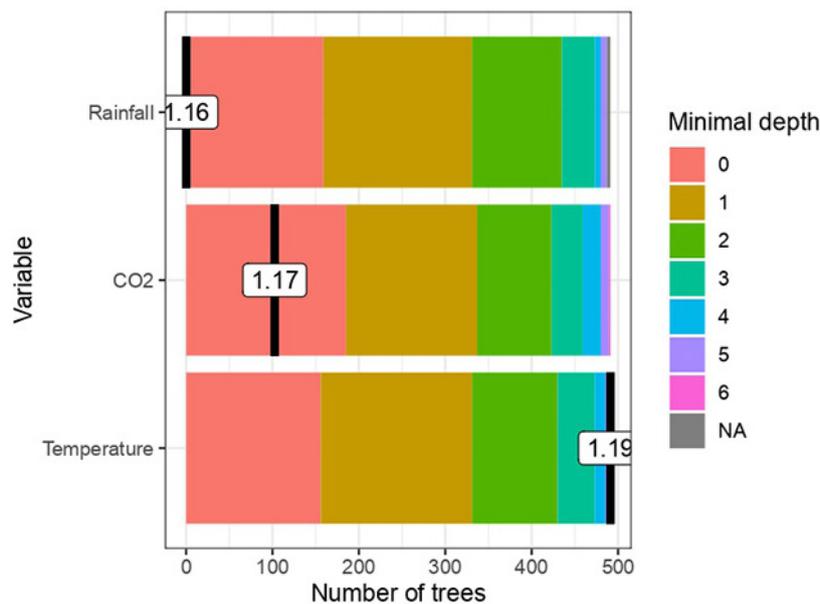

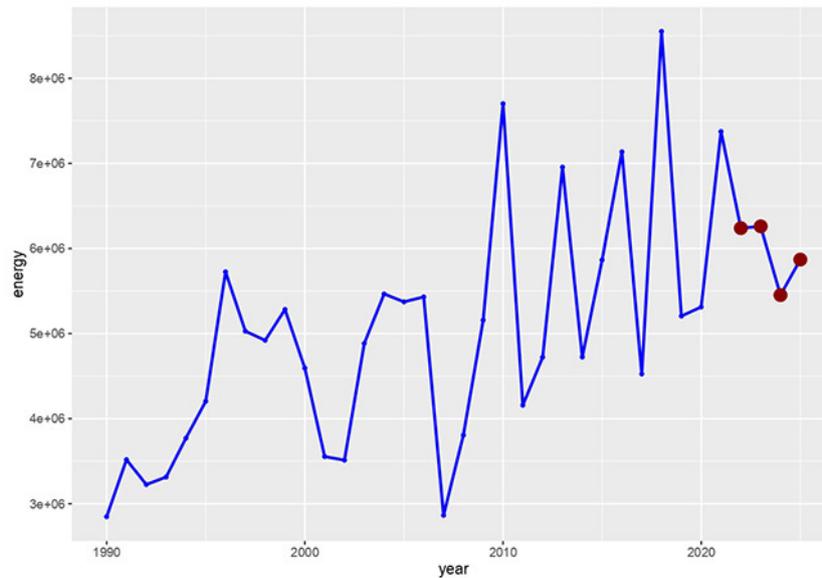**Figure 12.** Distribution of minimal depth

Figure 13. Energy generation by hydropower forecast

CONCLUSIONS

This study explored the climate change in Albania and its impact on energy production. Albania has experienced an increase in mean annual temperature of 1.5°C since the 1970s. Intensity, length, and frequency of heat waves are predicted to rise in the eastern Mediterranean, maybe up to six to eight times per year. Despite the considerable intra-seasonal variability of Albania's precipitation, a modest decline in mean annual precipitation has been noted since the 1970s. The rainfall in February, September, October, and December has decreased during the past ten years.

Several different models have been built for the study and prediction of temperature, rainfall and CO_2 emissions. The results obtained showed that Prophet algorithm performs better than other models for temperature and rainfall, while Prophet Boost algorithm is the best model for CO_2 emission prediction. The average annual temperature for the 4 years, 2022–2025 and the CO_2 emission will continue to increase. The predicted values received for each of the variables were placed in the best chosen model, in order to obtain the annual forecast of energy production by hydropower, for 2022 to 2025. Compared to 2021, energy production from hydropower will have a significant decrease in 2022, as well as in the following years.

Accuracy in these types of forecasts is not high, as extreme weather can affect the factors that impact energy production by hydropower. Moreover, consideration by governments of other renewable sources can make the energy produced by hydropower stay at normal capacities. One of them can be the energy produced by photovoltaic sources.

REFERENCES

1. Box G.E.P., Jenkins G.M. 1976. Time series analysis: forecasting and control. Revised Edition, Holden-Day, San Francisco, CA.

2. Breiman L. 2001. Random Forests. Machine Learning, 45, 5–32. DOI: 10.1023/A:1010933404324

3. Emodi N.V., Chaiechi T., Alam Beg A.R. 2018. The impact of climate change on electricity demand in Australia. Energy & Environment, 29(7), 1263–1297. DOI: 10.1177/0958305X18776538

4. Eysenbach J., Franklin B., Larsen A.J., Lindsey J. 2021. Predicting Power Using Time Series Analysis of Power Generation and Consumption in Texas. SMU Data Science Review, 5(3). https://scholar.smu.edu/datasciencereview/vol5/iss3/5

5. Franco G., Sanstad A.H. 2008. Climate change and electricity demand in California. Climatic Change, 87(1), 139–151, https://doi.org/10.1007/s10584-007-9364-y

6. Friedma J., Hastie T., Tibshirani R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1–22 https://doi.org/10.18637/jss.v033.i01

7. Gjika E., Ferrja A., Kamberi A. 2019. A study on the efficiency of hybrid models in forecasting precipitations and water inflow Albania case study. Advances in Science, Technology and Engineering Systems, 4(1), 302–310. DOI: 10.25046/aj040129

8. Gjika E., Basha L., Ferrja A., Kamberi A. 2021. Analyzing Seasonality in Hydropower Plants Energy Production and External Variables. Eng. Proc., 5(15). DOI: 10.3390/engproc2021005015

9. Gjika E., Basha L. 2022. A Comparative Study of Statistical and Deep Learning Models Used in Energy Load Prediction. World Conference on Sustainability, Energy and Environment. Paris, France.

10. Guo L.N., She C., Kong D.B., Yan S.L., Xu Y.P., Khayatnezhad M., Gholinia F. 2021. Prediction of the effects of climate change on hydroelectric generation, electricity demand, and emissions of greenhouse gases under climatic scenarios and optimized ANN model, Energy Reports, 7, 5431–5445. DOI: 10.1016/j.egyr.2021.08.134

11. Hale J., Long S. 2021. A Time Series Sustainability Assessment of a Partial Energy Portfolio Transition. Energies, 14(1), 141. DOI: 10.3390/en14010141

12. Harris R., Sollis R. 2003. Applied Time Series Modelling and Forecasting. Hoboken, NJ: John Wiley and Sons.

13. Hyndman R.J., Koehler A.B. 2006. Another look at measures of forecast accuracy, International Journal of Forecasting, 22(4), 679–688. DOI: 10.1016/j.ijforecast.2006.03.001.

14. Lahouar A., Ben Hadj Slama J. 2015. Random forests model for one day ahead load forecasting. IREC2015 The Sixth International Renewable Energy Congress, 1–6, DOI: 10.1109/IREC.2015.7110975

15. Mason K., Duggan J., Howley E. 2018. Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks, Energy, 155, 705–720. DOI: 10.1016/j.energy.2018.04.192

16. Meng M., Song C. 2020. Daily Photovoltaic Power Generation Forecasting Model Based on Random Forest Algorithm for North China in Winter. Sustainability, 12(6), 2247. DOI: 10.3390/su12062247

17. Serras P., Ibarra-Berastegi G., Sáenz J., Ulazia A. 2019. Combining random forests and physics-based models to forecast the electricity generated by ocean waves: A case study of the Mutriku wave farm, Ocean Engineering, 189, 106314. DOI: 10.1016/j.oceaneng.2019.106314

18. Tay J.K., Narasimhan B., Hastie T. 2021. Elastic Net Regularization Paths for All Generalized Linear Models. DOI: 10.48550/arXiv.2103.03475

19. Taylor S.J., Letham B. 2018. Forecasting at scale. The American Statistician, 72(1), 37–45, DOI: 10.1080/00031305.2017.1380080

20. Wadsack K., Acker T.L. 2019. Climate change and future power systems: the importance of energy storage in reduced-hydropower systems in the American Southwest, Clean Energy, 3(4), 241–250. DOI: 10.1093/ce/zkz018

21. Zhang C., Liao H., Mi Z. 2019. Climate impacts: temperature and electricity consumption. Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards, Springer. International Society for the Prevention and Mitigation of Natural Hazards, 99(3), 1259–1275, https://doi.org/10.1007/s11069-019-03653-w

22. The World Bank, https//www.worldbank.org