Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4913824, 12 pages
https://doi.org/10.1155/2021/4913824

Research Article

Regression-Based Prediction of Power Generation at Samanalawewa Hydropower Plant in Sri Lanka Using Machine Learning

Piyal Ekanayake,[1] **Lasini Wickramasinghe**,[1] **J. M. Jeevani W. Jayasinghe**,[1] **and Upaka Rathnayake**[2]

[1] *Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka*
[2] *Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka*

Correspondence should be addressed to J. M. Jeevani W. Jayasinghe; jeevani@wyb.ac.lk

Received 9 May 2021; Revised 19 June 2021; Accepted 26 July 2021; Published 31 July 2021

Academic Editor: Tzung-Pei Hong

This paper presents the development of models for the prediction of power generation at the Samanalawewa hydropower plant, which is one of the major power stations in Sri Lanka. Four regression-based machine learning and statistical techniques were applied to develop the prediction models. Rainfall data at six locations in the catchment area of the Samanalawewa reservoir from 1993 to 2019 were used as the main input variables. The minimum and maximum temperature and evaporation at the reservoir site were also incorporated. The collinearities between the variables were investigated in terms of Pearson's and Spearman's correlation coefficients. It was found that rainfall at one location is less impactful on power generation, while that at other locations are highly correlated with each other. Prediction models based on monthly and quarterly data were developed, and their performance was evaluated in terms of the correlation coefficient (R), mean absolute percentage error (MAPE), ratio of the root mean square error (RMSE) to the standard deviation of measured data (RSR), BIAS, and the Nash number. Of the Gaussian process regression (GPR), support vector regression (SVR), multiple linear regression (MLR), and power regression (PR), the machine learning techniques (GPR and SVR) produced the comparably accurate prediction models. Being the most accurate prediction model, the GPR produced the best correlation coefficient closer to 1 with a very less error. This model could be used in predicting the hydropower generation at the Samanalawewa power station using the rainfall forecast.

1. Introduction

Hydropower is one of the most widely used green energy sources in the world today. It is not only renewable but also highly reliable in generating and supplying power to national grids. Usually, major hydropower plants are used to generate electricity for the peak requirement of the countries. Most importantly, hydropower can be generated at a relatively low cost compared to other sources like thermal power. Therefore, there is an extensive demand for hydropower development in today's world. For example, Norway produces more than 95% of its energy requirement by hydropower while many other countries such as China, United States, Brazil, and Canada also produce more and more hydropower to meet their energy demands. This is mainly to achieve sustainable energy generation goals defined by the countries themselves.

Hydropower in Sri Lanka also plays an important role as the country now depends largely on thermal power generated by using imported coal and fuel oil. Sri Lanka was successful in generating green energy in the 1990s, but not much progress could be made due to the sudden increase in demand. Recent statistics indicate that Sri Lanka has produced an average of one-fifth of its energy demand from hydropower sources. Even though Sri Lanka has planned to enhance the generation of renewable power, there is little room for the construction of new major hydropower plants beyond the existing network of power stations. Out of the four types of hydropower development, viz., run-of-river, storage, pumped storage, and offshore hydropower plants, the first two types are very common in Sri Lanka, but the other types are still under discussion. Among the hydropower plants of storage type, the Samanalawewa hydropower development scheme showcases some important features due to its location (located in Sabaragamuwa province) and the relative high capacity for power generation. This hydropower plant is in the water rich Walawe basin, and the reservoir draws much attention not only from the perspective of the hydropower development but also due to its capacity as a primary source for irrigational purposes. Moreover, the hydropower scheme at Samanalawewa has drawn much attention due to a seepage leak from the reservoir. In this context, identifying the impact of climate change on the water resources is highly important for the Samanalawewa hydropower plant. Though a couple of studies addressed this problem recently, a comprehensive research on the prediction of power generation based on all related weather indices has not yet been conducted [1].

In a nonparametric statistical analysis of the monthly data over 26 years of the catchment rainfall associated with the Samanalawewa power plant in Sri Lanka, Dabare et al. [1] showed a positive correlation between the rainfall and the hydropower generation. While proposing nonlinear analysis for more specific conclusions, this study disavowed concerns on the negative impact of climate change on the rainfall. However, Suleiman and Ifabiyi [2] have revealed that the reservoir variables of inflow, storage, and the turbine release are strongly and positively correlated with the rainfall by analyzing the rainfall data around the Shiroro hydropower dam in Nigeria since 1990. Furthermore, they reported that the optimized turbine releases ensured the year-round power generation by the reservoir storage. However, a study on the impact of rainfall and temperature on electricity generation in Ghana pointed out that instability in climate dependent hydrology could cause uncertainties in hydropower generation [3].

Artificial neural network (ANN) was widely used to develop hydropower prediction models. Khaniya et al. [4] applied seven training algorithms in the ANN technique to predict the future power generation from 2020 to 2050 at the Samanalawewa hydropower plant in Sri Lanka using rainfall data for training and validation. Of these seven algorithms, the Quasi Newton algorithm outperformed the others in forecasting the hydropower to be generated within the next three decades for two climatic scenarios. This research further pointed out that other reservoir variables such as air temperature and humidity could also be used at the input layer of the model along with the other variables, such as reservoir inflow storage, turbine release, etc., which affect energy generation. A futuristic study was carried out to assess the impact of climate change on hydropower generation in Iran for two 3-decade periods (2020–2049 and 2070–2099) based on two climatic scenarios predicted by a regional climate model, in which the rainfall and hydropower generation were simulated by an ANN and a reservoir model [5]. This study found a positive impact of climate change on hydropower generation whose greater increase occurred during the first 3-decade period than the second. However, Beheshti et al. [5] expressed reservations on the uncertainties in predicting reservoir variables and hydropower under climate scenarios and suggested further studies, taking the variability in water allocation for irrigation into account. In addition, the complex nonlinear relationship between the rainfall and minihydropower generation in gauged and ungauged catchments of Sri Lanka has been studied recently using ANN, which showed a good correlation between them at the gauged catchments compared to ungauged catchments [6]. Based on the correlation values between the observed and predicted energies, Abdulkadir et al. [7] justified the use of neural network approaches in modelling the hydropower generation as a function of reservoir variables at two reservoirs along the River Niger in Nigeria. Developing predictive models of the hydropower generation in the Amazon, Lopes et al. [8] presented a comparative analysis between polynomial and ANNs using rainfall as the only input. Using three algorithms, group method of data handling (GMDH), ANN with Levenberg–Marquardt (ANN-LM), and ANN with Bayesian regulation (ANN-BR), it was shown that GMDH is the most appropriate algorithm to optimize the model result because of its adroitness in selecting the variables at the model entry layer and that ANN-LM algorithm failed to live up to expectations due to largely dispersed data and less accuracy.

Boadi and Owusu [9] used regression analysis to quantify the fluctuations in hydropower generation at the Akosombo hydroelectric power station in Ghana and emphasized the urgency in exploring alternative power sources to overcome energy security issues for sustainable development. Having used data over two consecutive 2-decade periods (1970–1990 and 1991–2010), their study reported that 21% of interannual fluctuation in power generation is accounted for by the rainfall variability, and that 72.4% of the same is explained by the El Niño-southern oscillation (ENSO) phenomenon and the lake water level. In another study, the streamflow and the potential hydropower generation were modelled using a data-based methodology in Mid Wales, where the projected impact of climate change on a hypothetical small power plant was assessed [10]. Its results showed an increase (decrease) in the streamflow and power output during winter (summer) months. Furthermore, Khaniya et al. [11] applied the Mann–Kendall test and Sen's slope estimator tests in a trend analysis to assess the performance of a minihydropower station in Sri Lanka based on 30-years of rainfall data and 6 years of electricity generation associated with the power plant. This study proved a positive rainfall trend at several rain gauging stations except in November and January while assuring the stability of the catchment area in the wake of climate variability. Nevertheless, research on regression-based prediction models to predict the hydropower generation in Sri Lanka is highly limited. Therefore, this research focuses on developing regression-based prediction models to predict the hydropower development capacity of the Samanalawewa hydropower development scheme.

In the next section on study area and data, the Samanalawewa catchment area, meteorological data used, and their relationship to power generation are elaborated.

Section 3 describes the regression techniques, methodology, and the evaluation criteria of the model performance. Section 4 presents the results and discussion where the models developed on monthly data, models based on quarterly rainfall data, and the salient features of the meteorological factors used are explained along with a comparison on findings from similar research work in some other countries. The paper is wrapped up with the major conclusions in Section 5.

2. Study Area and Data

2.1. Samanalawewa Catchment Area. Samanalawewa hydropower plant and its reservoir are in the Balangoda area in the Ratnapura district of Sri Lanka (coordinates of the power plant are 06°40′48″N, 80°47′54″E). The construction of the dam commenced in 1986 and commissioned in 1992. The project was carried out with financial support from Japan and the United Kingdom. It is a major hydropower scheme in the country and based on the Walawe River. The dam has a height of 110 m from its foundation and is 530 m in length. It is a rock-filled dam and holds 218 million m^3 of water out of 278 million m^3 of total capacity. The balance 60 million m^3 is kept for the dead storage [12].

The catchment area of the Samanalawewa reservoir is presented in Figure 1. The catchment area is around 372 km^2 and lies in the wet zone of the country, which receives a significant annual rainfall (annual average of 2867 mm) [13]. Therefore, the reservoir has a good overall water capacity throughout the year and generates 124 MW of electricity using two turbines.

A seepage leak was identified in the reservoir while it was under construction. Though it was treated at that time, the leakage continued even after the construction. As this is not through the dam, it has not caused any instability to the dam. The seepage is measured to be 2 m^3/s and, thereafter, that lost water is used to run a minihydropower station. For this reason, the Samanalawewa reservoir and the dam have captured the interest of power engineers. Due to all these reasons, it is highly important to analyze the hydropower scheme in light of changing climate and to forecast the power generation using key reservoir variables.

2.2. Meteorological Data and Their Observational Relationships to Power Generation

2.2.1. Rainfall Data. Twenty-six (26) years of rainfall data from 1993 to 2019 measured at 6 locations in the catchment area, Alupola, Detanagalla, Balangoda, Nagarak Estate, Belihuloya, and Nanperial, were purchased from the Department of Meteorology, the state repository of climate data in Sri Lanka. The highest mean annual rainfall during this period (4272 mm) was recorded at Alupola and the lowest (2170 mm) at Balangoda, while the other locations of Detanagalla, Nagarak Estate, Belihuloya, and Nanperial had received 2843 mm, 2247 mm, 2785 mm, and 2330 mm of annual mean rainfall, respectively. Table 1 shows the summary of major statistics (minimum, maximum, average, and standard deviation) of the monthly rainfall data at the six locations.

Coherent with the above mean annual figures, the highest and the lowest monthly average rainfalls (358 mm and 183 mm) are also reported from Alupola and Balangoda, respectively. The minimum values indicate that three locations (Nagarak Estate, Belihuloya, and Nanperial) have received no rainfall (0 mm) during the months mentioned in Table 1, while Detanagalla has experienced the highest monthly rainfall (1371 mm) in November 2006.

Figure 2 shows the monthly rainfall averaged over the period, 1993–2019, at the six locations in the catchment area. It can be seen that heavy rainfall has prevailed at each location during the months of April and November, which fall within the South-west and North-east monsoon periods of the country, respectively, and the slightly higher values in November imply the greater effect of the North-east monsoon than the South-west monsoon on the rainfall in the catchment area. It is also obvious that except at one location (Alupola), the least rainfall (upto 100 mm) has occurred during the 4-month period from June to September, which is less than one-third of the heavy rainfall during the monsoon periods.

Except during the 3 months from December to February, the solitary location of Alupola has continued to receive much higher rainfall producing the highest mean annual and the highest monthly average noticed in Table 1.

2.2.2. Evaporation Data. Figure 3 shows the monthly mean evaporation at the Samanalawewa reservoir site during the period from 1993 to 2019. According to this figure, the highest monthly mean evaporation (>4.5 mm) occurs during the 4-month period from June to September, which coincides with the same period with the least monthly rainfall averaged over the period of data at five locations described in Figure 2. The period from November to January indicates the lowest mean evaporation (<3.45 mm), while the monthly mean evaporation from February to October is greater than 4 mm. It can also be traced that subdued mean evaporation in April and November correspond to the monthly rainfall averages peaked in the same months, as shown in Figure 3.

2.2.3. Temperature Data. Figure 4 depicts the monthly mean maximum and minimum temperatures with their maxima and minima at the reservoir site for the period of 1993–2019. The lowest maximum temperature prevails during the cooler months of November to January, which picks up in February and maximizes in March and April. After the cooler months, the maximum temperature hovers between 33.8°C and 34.0°C and remains approximately the same (34–34.2°C) through the warmer months of July to September. Similarly, the minimum temperature reaches its lowest figures during the same cooler months but attains the highest values within 23.8°C to 24.4°C during June to August period. It picks up steadily from January to June and decreases gradually towards the cooler months.

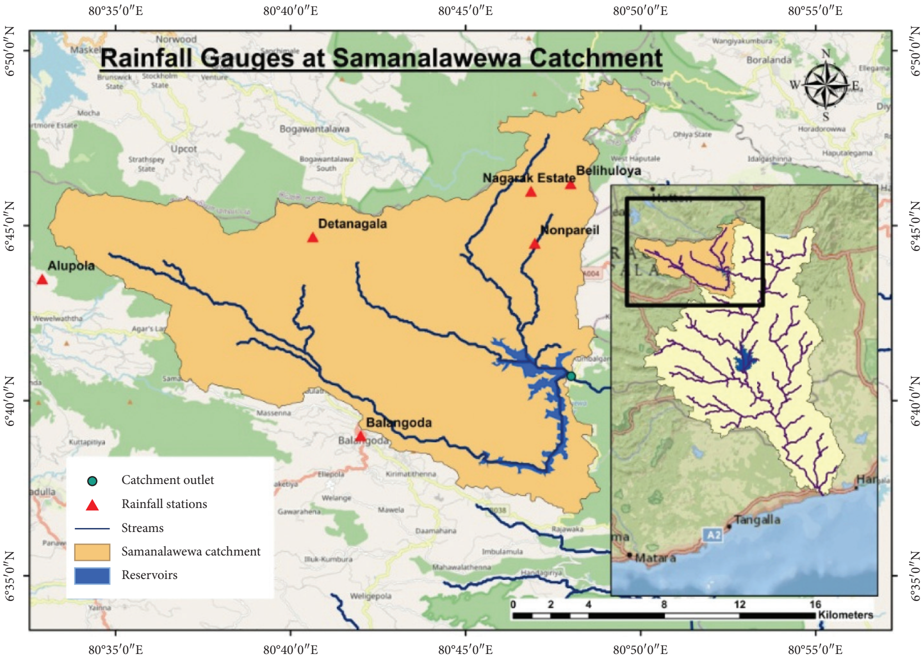


Figure 1: Catchment area of Samanalawewa reservoir.

Table 1: Summary of monthly rainfall data.

Location	Alupola (RF_1)	Detanagalla (RF_2)	Balangoda (RF_3)	Nagarak Estate (RF_4)	Belihuloya (RF_5)	Nanperial (RF_6)
Minimum rainfall (mm) and month occurred	24.5 12/1996	2.7 09/2016	4.7 05/1996	0.0 01/2009 02/2009 06/2012	0.0 09/2016	0.0 08/2001 07/2002
Maximum rainfall (mm) and month occurred	1160 05/2016	1371 11/2006	735 04/2015	661 11/2012	926 11/2012	930 11/2006
Average rainfall (mm)	358	239	183	188	233	193
Standard deviation (mm)	202	189	146	153	211	181

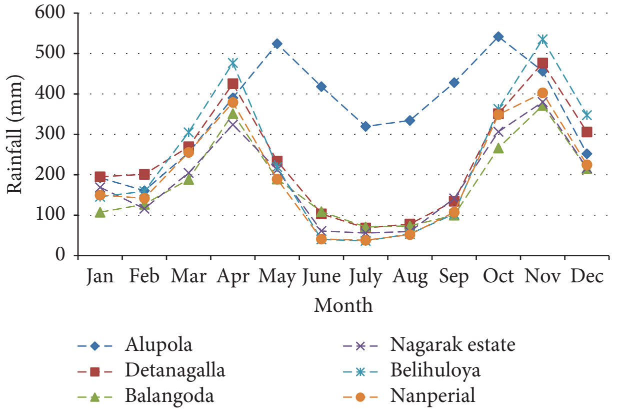


Figure 2: Monthly rainfall averaged over the period from 1993 to 2016.

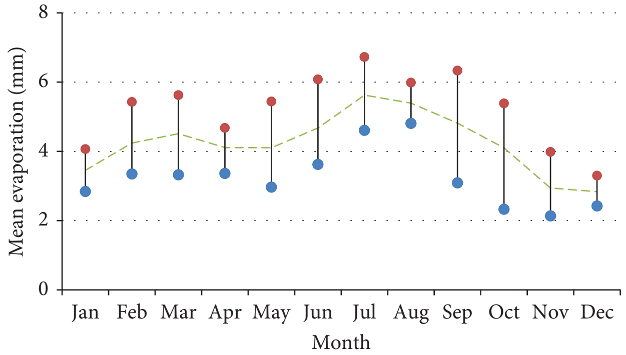


Figure 3: Monthly mean evaporation at the Samanalawewa reservoir site.

2.2.4. Power Generation Data. The annual power generation and its variation (from year 1993 – 2019) can be clearly seen from Figure 5. It can be traced that power generation has dropped sharply to 152 GWh in 1996, and since then, similar declines have occurred after every 5-6 years in 2002, 2007, 2012, and 2017 compared to the years around them. Similarly, the power generation has shown local maxima after every 5 years since 1993, and these maxima have occurred immediately after the years with local minima except in 1998.

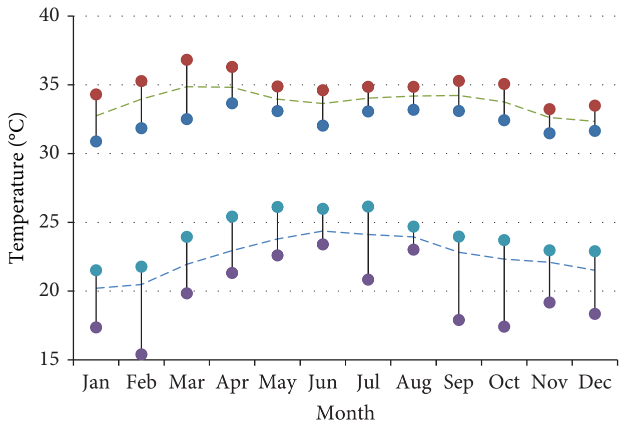


Figure 4: Monthly mean maximum and minimum temperatures.

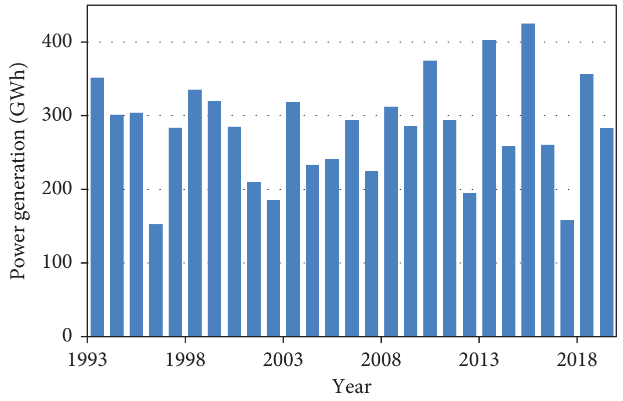


Figure 5: Annual power generation at the Samanalawewa power plant.

The minimum power generation (1.1 GWh) during the whole study period was found in November 2016. The associated rainfall during the preceding months of the same year was compiled with its November values at the six locations in Table 2 along with the power generated. This table shows that, except at Alupola, the rainfall has drastically decreased at other locations during the 4 month period from June to September, before picking up in October and reaching much higher values in November. Although power was generated uninterrupted, the effect of low rainfall has reflected through the nominal power outputs during September to November. It can also be understood that the rainfall experienced in November is comparable with the corresponding average values at each location presented in Figure 1, and that it has not created any positive impact on the power generated during the same month at the Samanalawewa power plant.

Furthermore, the monthly power generation averaged over the study period (1993–2019) was considered along with its maximum and minimum, shown in Figure 6. A detailed examination into data revealed that the maximum monthly power generation of 80.7 GWh was reported in January 1998, subsequent to a much higher rainfall since September 1997, e.g., a monthly rainfall over 340 mm at Alupola and Detanagalla. Moreover, Figure 6 shows the highest power generation during the two periods: April-May and November-January, which fall soon after the two months with the heaviest rainfall, April and November, indicated by the peaks in Figure 2. Therefore, it is evident that the rainfall of a particular month does not affect the power generation of the same month at Samanalawewa, which can justify the use of quarterly rainfall data for modelling instead of monthly data in this research.

3. Regression Techniques and Methodology

The hydropower generation at Samanalawewa from the year 1993 to 2019 was modelled in two time scales of monthly and quarterly data. Regression-based models were first developed by applying Gaussian process regression (GPR), support vector regression (SVR), multiple linear regression (MLR), and power regression (PR) to express the hydropower as a function of the catchment rainfall in monthly and quarterly scales. Then, another set of models was developed by applying the same techniques on multiple weather indices, viz., rainfall, mean reservoir evaporation, and mean minimum and maximum reservoir temperatures. Three options were considered based on the formation of quarterly data, such that Option 1 comprises of the grouping of months: Jan-Mar, Apr-Jun, Jul-Sep, and Oct-Dec, while Option 2 comprises of Feb-Apr, May-Jul, Aug-Oct, and Nov-Jan grouping. Option 3 included the clustering of months: Mar-May, Jun-Aug, Sep-Nov, and Dec-Feb. The models developed were then tested using the performance indicators given in equations (8)–(12) to understand the performance of the regression models.

The machine learning based models (SVR and GPR) were developed in the MATLAB environment (version 9.4.0.813654-R2018a), while the statistical models (MLR and PR) were developed by programming in the *R* software (R 4.0.3).

3.1. Support Vector Regression.

3.1. Support Vector Regression. Support vector regressions (SVRs) are supervised machine learning models based on a regression algorithm that can deal with nonlinear data for prediction. It is highlighted due to its robustness and high prediction accuracy in the presence of dimensionality of the input space [14]. The training and testing data used in SVR are assumed to be independent and identically distributed having an unknown probability function. SVR develops a linear hyperplane that transforms multidimensional input vectors (weather indices) into output values (power generation), which are then used to predict future output values. For linear function f, a set of n number of data points $P = (x_i, y_i)$, where x_i is the input vector of a data point i and y_i is its actual value, the hyperplane $f(x)$ is given as follows [15]:

$$f(x) = wx_i + b, \tag{1}$$

where w is the slope and b is the intercept. For nonlinear relations, a map φ that translates x_i into a higher-dimensional feature space needs is defined. Then, w becomes a function of $\varphi(x_i)$, and the Kernel function is defined as a product as follows:

Table 2: Power generation and the rainfall received from June to November of 2016.

Month	Power (GWh)	Rainfall (mm)					
		Alupola (RF_1)	Detanagalla (RF_2)	Balangoda (RF_3)	Nagarak Estate (RF_4)	Belihuloya (RF_5)	Nanperial (RF_6)
Jun 2016	33.4	236	26	61	18	8	18
Jul 2016	13.5	148	37	29	39	49	40
Aug 2016	13.4	241	8	19	14	19	14
Sep 2016	2.9	220	3	17	6	0	6
Oct 2016	2.5	581	93	131	100	77	102
Nov 2016	1.1	591	350	419	349	518	358

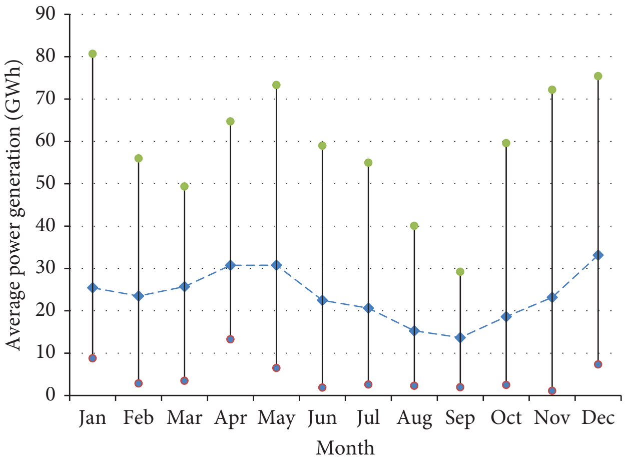


Figure 6: Monthly power generation averaged over the period from 1993 to 2019.

$$k(x_i, x) = \phi(x_i)\phi(x). \qquad (2)$$

In this research, 5-fold crossvalidation was applied using 4 folds for training and the other fold for evaluation. It was repeated 5 times, using one different fold for evaluation each time. SVR-based prediction models were developed by applying Kernel functions of linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian, and the model that gives the lowest RMSE was selected for further analysis.

3.2. Gaussian Process Regression. Gaussian distribution is defined by its mean and the standard deviation, characterized by a symmetrical curve about the mean that coincides with the mode and the median. In statistical analysis, a Gaussian process is a stochastic process with every finite collection of random variables having a multivariate normal distribution [16]. Gaussian process regression (GPR) is nonparametric and useful in dealing with small datasets. Another advantage is its capacity to address uncertainty measurements of the predictions. A Gaussian process is denoted as follows [17]:

$$f(x) = \mathrm{GP}(m(x), k(x, \acute{x})), \qquad (3)$$

where $m(x)$ and $k(x, \acute{x})$ are the mean function and the covariance function, respectively. The mean function $m(x)$ is the expectation of the function $f(x)$ at the point x, and the covariance function is a measure of the confidence level for $m(x)$. In this research, GPR-based models were developed by applying Kernel functions of rational quadratic, exponential, squared exponential, and Matern 5/2, and the model with the lowest RMSE was selected for further analysis.

3.3. Multiple Linear Regression. Multiple linear regression (MLR) assumes a linear relationship among the independent and dependent variables. Therefore, the best fit is described by a straight line of the relationship wherein the data are assumed to be normally distributed [18]. The general mathematical formula of the MLR model for n number of independent variables is written as follows [19]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i + \cdots + \beta_n x_n + \varepsilon, \qquad (4)$$

where y is the dependent variable (power generation), β_0 is the intercept on the y axis, β_i is the slope coefficient of the i^{th} input variable x_i, and ε is the model error.

3.4. Power Regression. Power regression (PR) develops a power relationship among the variables. The nonlinearity of data was considered in PR, which modelled the power generation proportional to the product of powers of the independent variables as follows [20–22]:

$$y = a x_1^b x_2^c \cdots x_n^p, \qquad (5)$$

where n is the number of observations and $a, b, c, \ldots, p$ are constants.

3.5. Correlation Coefficient. Pearson's and Spearman's correlation coefficients were used to assess the collinearity among each pair of input and output variables. Monthly and quarterly data were used to determine the correlation coefficient.

Pearson's correlation coefficient is the most commonly used test statistic for measuring the linear dependency of two normally distributed random variables as it takes both variance and the covariance into account [23]. It indicates both the degree and the direction of the association, if any. Pearson's correlation coefficient (R_p) of two random variables X and Y is mathematically presented as follows [24]:

$$R_P = \frac{\text{covariance}\,(X,Y)}{\sqrt{\text{variance}\,(X)}\sqrt{\text{variance}\,(Y)}} = \frac{\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{N}\,(x_i-\overline{x})^2\sum_{i=1}^{N}\,(y_i-\overline{y})^2}}, \tag{6}$$

where $-1 \leq R_P \leq +1$. The values of R_P closer to ±1 are the evidence for strong associations, which should be reflected on the scatter plot between the two variables with close congregation of points around the line of the best fit. The intervals [±0.66, ±1], [±0.33, ±0.65], and [±0.32, 0] of R_P are considered as strong, medium, and low degree correlations, respectively.

Spearman's correlation coefficient may be viewed as the nonparametric counterpart of Pearson's correlation coefficient for nonlinear data, which also measures both the strength and direction of the two variables [25]. Its value also varies between −1 and +1 having a similar interpretation as for Pearson's correlation coefficient. The mathematical form of Spearman's correlation coefficient (r_s) is defined as follows when it is applied to n pairs of rank variables, and the ranks are distinct integers,

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)}, \tag{7}$$

where d_i is the difference between the ranks of the two observations.

3.6. Evaluation Criteria of Developed Models. The following statistical measures: the correlation coefficient (R), mean absolute percentage error (MAPE), ratio of the root mean square error (RMSE) to the standard deviation of the measured data (RSR), BIAS, and the Nash number were used to evaluate the dexterity of each model developed in the present study based on the mathematical formula indicated in the following equations:

$$\text{correlation coefficient; } R = \frac{\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{N}\,(x_i-\overline{x})^2\sum_{i=1}^{N}\,(y_i-\overline{y})^2}}, \tag{8}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{x_i-y_i}{x_i}\right| \times 100, \tag{9}$$

$$\text{RSR} = \frac{\sqrt{\text{MSE}}}{\sigma_x}, \tag{10}$$

$$\text{BIAS} = \frac{\sum_{i=1}^{N}(y_i-x_i)}{N}, \tag{11}$$

$$\text{Nash number} = 1 - \left[\frac{\sum_{i-1}^{N}\,(x_i-y_i)^2}{\sum_{i-1}^{N}\,(x_i-\overline{x})^2}\right], \tag{12}$$

where x_i is the actual power generation, y_i is the predicted power generation, $\overline{x}$ and $\overline{y}$ are their means, N is the number of data values, and σ_x is the standard deviation of actual power generation. The values of MAPE and RSR closer to zero and R the Nash number closer to 1 imply more accurate models for the prediction of power generation. A zero BIAS means accurate models, whereas its negative and positive values would indicate underestimation and overestimation, respectively.

4. Results and Discussion

The following subsections present the results obtained from the regression analysis for hydropower generation at the Samanalawewa hydropower plant based on the catchment rainfall, reservoir evaporation, and temperature. The analysis was carried out using the regression models described in the previous section.

4.1. Models Developed Based on Monthly Data. Correlations between the hydropower generation and the monthly rainfall of six rain gauges in the catchment area are presented in Table 3. Results clearly show that there is very little correlation between the power generation and rainfall at monthly scale.

This observation is further consolidated by the performance (R) of the regression models in the monthly scale, shown in Table 4. Out of the SVR models developed by applying six types of kernels, the fine Gaussian SVR demonstrated the best performance. Exponential GPR is the most accurate among the GPR models developed by applying four kernels. The results revealed that none of the regression-based prediction models is accurate when the monthly rainfall at the catchment area is used as the input variables.

Based on these results, it can be clearly concluded that the monthly scale is not appropriate for regression analysis in compliance with the observations drawn from Table 2. Therefore, quarterly models were developed by using quarterly rainfall data as input variables.

4.2. Quarterly Models Developed Based on Rainfall Data. The following results presented in Table 5 and Figure 7 are based on the models developed with respect to the quarterly rainfall data. Figure 7 shows the relationship between the observed power generation and the predicted power generation produced by the regression-based prediction models.

Based on the deviations of the predictions, it can be clearly seen that the machine learning models (Figures 7(a) and 7(b)) outperform the statistical models (Figures 7(c) and 7(d)). Fine Gaussian SVR outperformed the other five types of SVR-based models, while the rational quadratic GPR was the most accurate among the GPR-based models. Power generation values predicted by the SVR and GPR models are closer to the reality, which correspond to the coefficient of correlation reaching 1 with least error in

Table 3: Coefficient of correlation between hydropower generation and monthly rainfall.

Rainfall of rain gauges	RF_1	RF_2	RF_3	RF_4	RF_5	RF_6
Coefficient of correlation	0.07	0.25	0.24	0.17	0.16	0.11

Table 4: Performance of the prediction models for monthly rainfall data.

Regression technique	SVR	GPR	MLR	PR
R	0.25	0.28	0.29	0.39

Table 5: Performance of the regression models based on quarterly rainfall.

Statistical measure (performance indicator)	Regression technique			
	SVR	GPR	MLR	PR
R	0.86	0.95	0.49	0.61
MAPE (%)	20.2	7.0	60.3	39.3
BIAS	−0.7	0.4	7.1	−4.9
Nash	0.7	0.9	0.2	0.2
RSR	0.5	0.3	0.9	0.9

terms of MAPE, BIAS, Nash number, and the RSR (Table 5). The excellence of GPR compared to SVR is evident from the highest *R* and Nash number, least MAPE and RSR, and a smaller BIAS.

Moreover, the coefficients of correlation are much higher in Table 5 compared to those in Table 4, which reinforces the appropriateness of using quarterly data instead of the monthly data. Among the four techniques, the models based on SVR and GPR show much better performance compared to the other two models. The MLR model has the lowest performance as indicated by the performance evaluators of *R* and the MAPE in particular. Furthermore, it has the highest BIAS and RSR as well. Therefore, compared to other regression models, the GPR model can be recommended as an outstanding technique.

4.3. Quarterly Models Developed Based on Four Meteorological Factors. Table 6 summarizes the correlation coefficients generated by all the models for the three seasonal options tested on quarterly basis with respect to the four climatic variables. In all three seasonal options, fine Gaussian SVR was the best among SVR-based models. Rational quadratic outperforms other GPR kernels in the first and second seasonal options, while Matern 5/2 was the best GPR in the third seasonal option. As was seen in Table 5, the GPR model has outperformed the other regression models. Furthermore, equally better performance can be seen between the GPR and SVR models. Similarly close results are observed between these models irrespective of the three seasonal clusters used in the quarterly analysis. In addition, the correlation coefficients suggest that the MLR and PR are not the best regression techniques to predict the hydropower generation in the Samanalawewa hydropower plant in Sri Lanka.

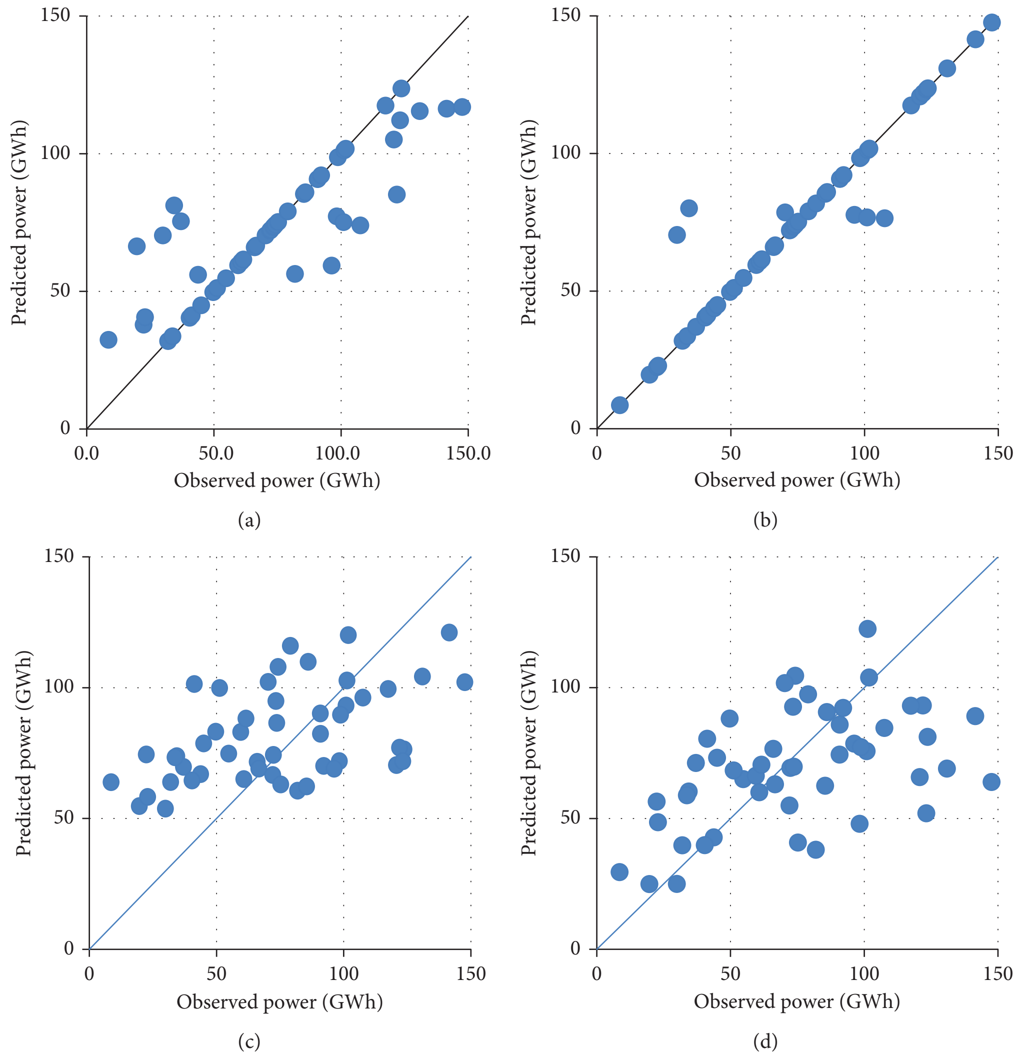
Table 7 presents Pearson's and Spearman's correlation coefficients between the power generation and each catchment rainfall of the six rain gauges and among the paired rain gauges. According to the interpretation of the size of these coefficients introduced in Section 3.5, it can be noticed that very strong pairwise correlations exist between the rainfall received in the catchment areas of Balangoda (RF_3), Nagarak Estate (RF_4), Belihuloya (RF_5), and Nanperial (RF_6), respectively. Moderate correlations appear between the power generation and each of the five rain gauges except at Alupola (RF_1). The only exception with the weakest correlation between rainfall and the power generation is reported from Alupola.

Figure 8 illustrates the relationship between the predicted and the observed power generation. The strong linear relationships between the observed and predicted values in Figures 8(a) and 8(b) indicate that machine learning (SVR and GPR) models forecast the hydropower generation with remarkable accuracy (more than 87%). However, the predicted power generation for the MLR and PR regression models is scattered around the line of best fit as shown in Figures 8(c) and 8(d).

The results shown in Table 6 and Figure 8 are further verified by the model performance indicators in Table 8, which arise from the four regression models applied for Option 1. The GPR regression model presents the best results with the lowest errors and the highest correlation coefficient. Therefore, it can be concluded that the GPR regression model is a better regression model compared to others to predict the hydropower generation in the Samanalawewa hydropower plant.

Similar observations and findings could be seen in the other two options too (which are not shown here). Therefore, the superiority of the GPR model could be generalized for the power generation at Samanalawewa irrespective of the seasonal options.

4.4. Comparison of Similar Research. Table 9 presents a summary of some related work in the literature on the prediction of hydropower generation based on climatic data and using different modelling techniques in several countries. Most of the research studies are based on ANNs. A major drawback of ANNs was discussed in the introduction section of this paper. Even though they showcased better results, the black box environment in analysis leads to less information of the relationship. Some other methods like stepwise regression have also been used to predict the hydropower development. However, in most of these studies, only one statistical measure, i.e., correlation coefficient, was used to evaluate the prediction accuracy. Therefore, it could be analytically proved with evidence that out of the four prediction models developed in this study, the GPR has shown excellent performance and even outperformed all the models cited here. In particular, in the previous study conducted on the Samanalawewa hydropower generation, only the ANN was applied, and the performance was evaluated only in terms of the correlation coefficient and the MSE [6]. All the ANN-based prediction models were found



(a)

(b)

(c)

(d)

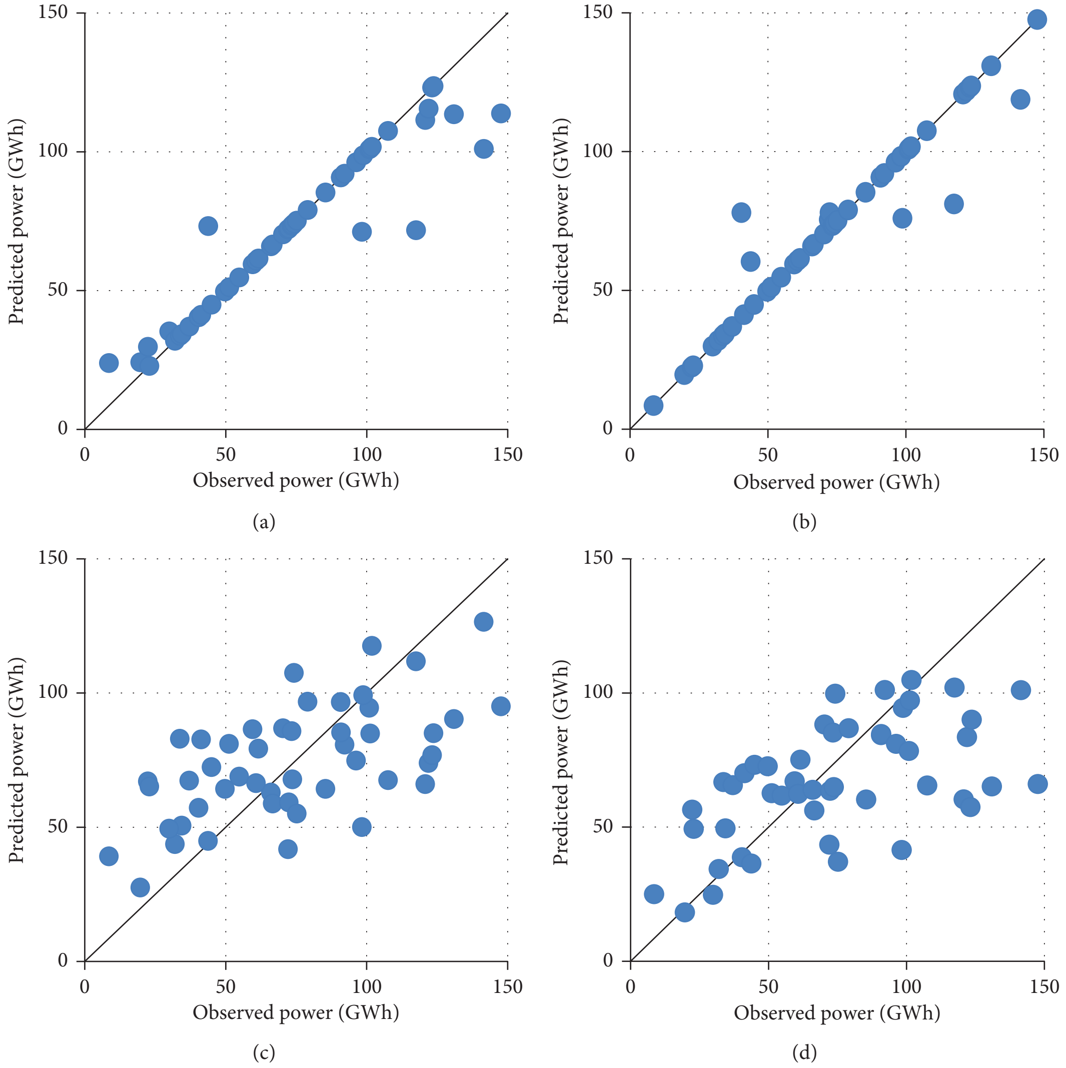
Figure 7: Predicted power generation against the observed power generation. (a) For SVR. (b) For GPR. (c) For MLR. (d) For PR.

Table 6: Correlation coefficients for the regression models based on quarterly climatic data.

	Regression technique			
	SVR	GPR	MLR	PR
Option 1	0.87	0.92	0.60	0.67
Option 2	0.87	0.91	0.44	0.45
Option 3	0.91	0.94	0.44	0.45

Table 7: Matrix of Pearson's (R) and Spearman's (r_s) correlation coefficients.

Power	1						
RF_1	$R = 0.10$ $r_s = 0.11$	1					
RF_2	$R = 0.35$ $r_s = 0.39$	$R = 0.38$ $r_s = 0.39$	1				
RF_3	$R = 0.33$ $r_s = 0.35$	$R = 0.50$ $r_s = 0.50$	$R = 0.83$ $r_s = 0.90$	1			
RF_4	$R = 0.45$ $r_s = 0.46$	$R = 0.36$ $r_s = 0.34$	$R = 0.85$ $r_s = 0.86$	$R = 0.85$ $r_s = 0.87$	1		
RF_5	$R = 0.35$ $r_s = 0.39$	$R = 0.38$ $r_s = 0.40$	$R = 0.90$ $r_s = 0.94$	$R = 0.94$ $r_s = 0.94$	$R = 0.90$ $r_s = 0.90$	1	
RF_6	$R = 0.34$ $r_s = 0.38$	$R = 0.28$ $r_s = 0.30$	$R = 0.88$ $r_s = 0.89$	$R = 0.80$ $r_s = 0.85$	$R = 0.90$ $r_s = 0.91$	$R = 0.88$ $r_s = 0.91$	1
	Power	RF_1	RF_2	RF_3	RF_4	RF_5	RF_6



(a)

(b)

(c)

(d)

Figure 8: Predicted power generation vs. the observed power generation. (a) For SVR, (b) For GPR, (c) For MLR, (d) For PR.

Table 8: Performance of the models based on quarterly climate data for option 1.

Statistical measure (performance indicator)	Regression technique			
	SVR	GPR	MLR	PR
R	0.87	0.92	0.60	0.67
MAPE (%)	9.7	4.5	46.1	35.7
BIAS	−2.5	−0.4	−0.01	−7.1
Nash	0.9	0.9	0.4	0.3
RSR	0.4	0.3	0.8	0.8

Table 9: Comparison of previous related studies.

Ref	Country of study	Input variables	Modeling technique	Performance of the models
[3]	Ghana	Temperature and rainfall	Statistical analysis	—
[6]	Sri Lanka	Rainfall	ANN (LM)	$R = 0.86$ $\text{MSE} = 1.03 \times 10^6$
			ANN (BR)	$R = 0.73$ $\text{MSE} = 8.9 \times 10^3$
			ANN (SCG)	$R = 0.76$ $\text{MSE} = 7.42 \times 10^5$

Table 9: Continued.

Ref	Country of study	Input variables	Modeling technique	Performance of the models
[7]	Nigeria	Evaporation losses, reservoir inflow, storage, reservoir elevation, turbine release, net generating head, plant use coefficient, tail race level	ANN	$R = 0.89$

[8]	Brazil	Rainfall at seven subbasins	Group method of data handling (GMDH)	$R = 0.90$
				MAE = 443
				MAPE = 12.34%

			ANN (BR)	$R = 0.88$
				MAE = 450
				MAPE = 12.41%

			ANN (LM)	$R = 0.83$
				MAE = 593
				MAPE = 17%

[9]	Ghana	Rainfall, ENSO, lake level elevation, and net lake inflow	Stepwise multiple regression	$R^2 = 0.753$ Adjusted $R^2 = 0\ .742$

less accurate than the GPR-based model presented in this paper. In this sense, the scientific contribution of the present paper is well justified.

5. Conclusions

The paper presented highly accurate models for the prediction of hydropower generation by using machine learning techniques. Particularly, the GPR-based prediction models outperformed the other techniques used in this research, as well as in similar studies conducted on hydropower plants located in other countries. Therefore, when the future rainfall of the catchment area is known by forecast, the power generation at the Samanalawewa hydropower station can be predicted accurately. It could also be concluded that the monthly rainfall is not reflected through the power generated during the same month at Samanalawewa. The lack of correlation between the hydropower generation and the monthly rainfall of rain gauges in the catchment area clearly indicated that monthly data are not the best for forecasting the power generation, rather it is the quarterly rainfall that produced the most accurate predictions with high correlation.

The prediction of power generation at this major power plant in Sri Lanka will certainly provide useful information, not only for the energy authorities of the country but also for the policy makers, investors, and the government in ensuring uninterrupted power supply through an environmentally friendly renewable source at affordable cost to the consumers. The climate models can effectively be used in forecasting the climate patterns for future years under different representative concentration pathways (*RCP2.6, RCP4.5, RCP6, and RCP8.5*). These predicted climate data can be used in the prediction models developed in this study to forecast the hydropower generation at the Samanalawewa hydropower plant in future years (in 2030 to 2099). Thus, the findings of this research would be highly useful for the future planning processes.

Data Availability

The climatic data and the analysis data are available from the corresponding author upon request.

Disclosure

The research was carried out at the Wayamba University of Sri Lanka and the Sri Lanka Institute of Information Technology environments.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the support from the Ceylon Electricity Board, Sri Lanka, and the support by Dr. Kamal Laksiri in providing the Samanalawewa Hydropower generation data. In addition, the authors would like to thank Ms. Imiya Chathurinika for creating the catchment drawing for the Samanalwewa catchment.

References

[1] G. Dabare, M. B. Gunathilake, N. Miguntanna, K. Laksiri, and U. Rathnayake, "Climate variation and hydropower generation in samanalawewa hydropower scheme, Sri Lanka," *Engineer: Journal of the Institution of Engineers, Sri Lanka*, vol. 53, no. 3, pp. 19–25, 2020.

[2] Y. Suleiman and L. Ifabiyi, "The role of rainfall variability in reservoir storage management at Shiroro Hydropower Dam, Nigeria," *Momona Ethiopian Journal of Science*, vol. 7, no. 1, pp. 55–63, 2015.

[3] A. Kabo-Bah, C. Diji, K. Nokoe, Y. Mulugetta, D. Obeng-Ofori, and K. Akpoti, "Multiyear rainfall and temperature trends in the Volta river basin and their potential impact on hydropower generation in Ghana," *Climate*, vol. 4, no. 4, p. 49, 2016.

[4] B. Khaniya, C. Karunanayake, M. B. Gunathilake, and U. Rathnayake, "Projection of future hydropower generation in Samanalawewa power plant, Sri Lanka," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8862067, 11 pages, 2020.

[5] M. Beheshti, A. Heidari, and B. Saghafian, "Susceptibility of hydropower generation to climate change: Karun III Dam case study," *Water*, vol. 11, no. 5, Article ID 1025, 2019.

[6] A. Perera and U. Rathnayake, “Relationships between hydropower generation and rainfall-gauged and ungauged catchments from Sri Lanka,” *Mathematical Problems in Engineering*, vol. 2020, Article ID 9650251, 8 pages, 2020.

[7] T. S. Abdulkadir, A. W. Salami, A. R. Anwar, and A. G. Kareem, “Modelling of hydropower reservoir variables for energy generation: neural network approach,” *Ethiopian Journal of Environmental Studies and Management*, vol. 6, no. 3, pp. 310–316, 2013.

[8] M. N. G. Lopes, B. R. P. da Rocha, A. C. Vieira, J. A. S. de Sá, P. A. M. Rolim, and A. G. da Silva, “Artificial neural networks approaches for predicting the potential for hydropower generation: A case study for Amazon region,” *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 6, pp. 5757–5772, 2019.

[9] S. A. Boadi and K. Owusu, “Impact of climate change and variability on hydropower in Ghana,” *African Geographical Review*, vol. 38, no. 1, pp. 19–31, 2019.

[10] D. Carless and P. G. Whitehead, “The potential impacts of climate change on hydropower generation in Mid Wales,” *Hydrology Research*, vol. 44, no. 3, pp. 495–505, 2013.

[11] B. Khaniya, H. G. Priyantha, N. Baduge, H. M. Azamathulla, and U. Rathnayake, “Impact of climate variability on hydropower generation: A case study from Sri Lanka,” *ISH Journal of Hydraulic Engineering*, vol. 26, no. 3, pp. 301–309, 2020.

[12] K. Laksiri, “A modern addition—uncovering Samanalawewa,” *International Water Power and Dam Construction*, Archived from the original on 2010-01-09. Retrieved 2009-06-27, 2004.

[13] K.-H. Nagel, “Limits of the geological predictions constructing the Samanalawewa pressure tunnel, Sri Lanka,” *Bulletin of the International Association of Engineering Geology*, Springer, vol. 45, no. 1, , pp. 97–110, Berlin, Germany, 1992.

[14] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Springer nature, Basingstoke, UK, 2015.

[15] M. Wauters and M. Vanhoucke, “Support vector machine regression for project control forecasting,” *Automation in Construction*, vol. 47, pp. 92–106, 2014.

[16] D. Kong, Y. Chen, and N. Li, “Gaussian process regression for tool wear prediction,” *Mechanical Systems and Signal Processing*, vol. 104, pp. 556–574, 2018.

[17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, MA, USA, 2006.

[18] G. K. Uyanık and N. Güler, “A study on multiple linear regression analysis,” *Procedia-Social and Behavioral Sciences*, vol. 106, pp. 234–240, 2013.

[19] M. Tranmer and M. Elliot, “Multiple linear regression,” *The Cathie Marsh Centre for Census and Survey Research (CCSR)*, vol. 5, no. 5, pp. 1–5, 2008.

[20] R. G. Shepherd, “Regression analysis of river profiles,” *The Journal Of Geology*, vol. 93, no. 3, pp. 377–384, 1985.

[21] S. d. Jong, B. M. Wise, and N. L. Ricker, “Canonical partial least squares and continuum power regression,” *Journal Of Chemometrics*, vol. 15, no. 2, pp. 85–100, 2000.

[22] V. Gowariker, V. Thapliyal, R. Sarker, G. Mandal, and D. Sikka, “Parametric and power regression models: New approach to long range forecasting of monsoon rainfall in India,” *Mausam*, vol. 40, no. 2, pp. 115–122, 1989.

[23] K. H. Zou, K. Tuncali, and S. G. Silverman, “Correlation and simple linear regression,” *Radiology*, vol. 227, no. 3, pp. 617–628, 2003.

[24] D.-S. Lee, C.-S. Chang, and H.-N. Chang, “Analyses of the clustering coefficient and the Pearson degree correlation coefficient of chung’s duplication model,” *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 3, pp. 117–131, 2016.

[25] J. H. Zar, “Spearman rank correlation,” *Encyclopedia of Biostatistics*, Wiley, vol. 7Hoboken, NJ, USA, , 2005.