water

MDPI

Article

Monthly Streamflow Prediction by Metaheuristic Regression Approaches Considering Satellite Precipitation Data

Mojtaba Mehraein [1,*], Aadhityaa Mohanavelu [2], Sujay Raghavendra Naganna [3], Christoph Kulls [4] and Ozgur Kisi [4,5,*]

[1] Faculty of Engineering, Kharazmi University, Tehran 15719-14911, Iran
[2] Department of Civil Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India
[3] Department of Civil Engineering, Siddaganga Institute of Technology, Tumakuru 572103, India
[4] Department of Civil Engineering, University of Applied Sciences, 23562 Lübeck, Germany
[5] Department of Civil Engineering, Ilia State University, 0162 Tbilisi, Georgia
* Correspondence: mehraein@khu.ac.ir (M.M.); ozgur.kisi@th-luebeck.de (O.K.)

Abstract: In this study, the viability of three metaheuristic regression techniques, CatBoost (CB), random forest (RF) and extreme gradient tree boosting (XGBoost, XGB), is investigated for the prediction of monthly streamflow considering satellite precipitation data. Monthly streamflow data from three measuring stations in Turkey and satellite rainfall data derived from Tropical Rainfall Measuring Mission (TRMM) were used as inputs to the models to predict 1 month ahead streamflow. Such predictions are crucial for decision-making in water resource planning and management associated with water allocations, water market planning, restricting water supply and managing drought. The outcomes of the metaheuristic regression methods were compared with those of artificial neural networks (ANN) and nonlinear regression (NLR). The effect of the periodicity component was also investigated by importing the month number of the streamflow data as input. In the first part of the study, the streamflow at each station was predicted using CB, RF, XGB, ANN and NLR methods and considering TRMM data. In the second part, streamflow at the downstream station was predicted using data from upstream stations. In both parts, the CB and XGB methods generally provided similar accuracy and performed superior to the RF, ANN and NLR methods. It was observed that the use of TRMM rainfall data and the periodicity component considerably improved the efficiency of the metaheuristic regression methods in modeling (prediction) streamflow. The use of TRMM data as inputs improved the root mean square error (RMSE) of CB, RF and XGB by 36%, 31% and 24%, respectively, on average, while the corresponding values were 37%, 18% and 43% after introducing periodicity information into the model's inputs.

Keywords: streamflow prediction; metaheuristic regression approaches; satellite precipitation data; TRMM

check for updates

Citation: Mehraein, M.; Mohanavelu, A.; Naganna, S.R.; Kulls, C.; Kisi, O. Monthly Streamflow Prediction by Metaheuristic Regression Approaches Considering Satellite Precipitation Data. *Water* **2022**, *14*, 3636. https://doi.org/10.3390/w14223636

Academic Editor: Paul Kucera

Received: 2 October 2022
Accepted: 9 November 2022
Published: 11 November 2022

1. Introduction

Streamflow is one of the most important components of the terrestrial water cycle. It describes the flow of water that enters the watershed as precipitation, reaching its destination through natural drainage into lakes and oceans by flowing through creeks, streams, and rivers [1,2]. While high streamflow in a channel (stream or river) can cause flooding and waterlogging, low streamflow can adversely affect dependent riverine ecosystems [3,4]. In both cases, the consequences can be dire, resulting in severe socio-economic losses and ecosystem fragmentation [5,6]. Hence, it is necessary to accurately predict streamflow so that a significant portion of these damages can be effectively mitigated. Further, since surface water reservoirs are principally used to supply water to satisfy urban drinking water requirements, accurate streamflow forecasting is required for the efficient planning and management of these systems [7,8]. Several variables, including the climate and hydrology, the hydraulic properties of the stream, elevation, and the presence of upstream controls, affect streamflow. As the uncertainty and hydro-climatic load on a stream and river network increase, it becomes increasingly difficult to precisely forecast streamflow, making it an arduous task to reduce or control its vulnerability.

Broadly speaking, streamflow prediction models can be classified as either linear or non-linear [9,10]. Traditional linear regression models, such as autoregressive integrated moving average models and multiple linear regression techniques, have been used for streamflow forecasting [11,12]. These models can perform well in forecasting long lead time streamflow forecasts, but their performance is limited by the assumption of linearity of the streamflow [10]. Since streamflow time series are inherently non-linear owing to their stochastic nature and dependency on different external control (exogenous) variables [13], non-linear modeling techniques such as artificial neural networks (ANN), support vector regression (SVM), extreme learning machine (ELM), and tree-based regression techniques such as random and rotation forest (RF) have been widely applied in streamflow forecasting [10]. Most of these models are basically machine learning models that leverage time series data to make accurate predictions. These models are also being successfully used to forecast other natural and hydro-geo-climatic phenomena [14–16].

The development and use of ANN and ANN ensemble (hybrid) models in streamflow forecasting have been documented in several studies [17–19]. In several case studies, the application of ELM, SVM, and RF models to understand the potential of these models in streamflow forecasting revealed that each showed varying performance under different hydro-climatic and geographic conditions [19,20]. Though all the above-mentioned models performed decently in forecasting streamflow, obtaining a higher correlation coefficient (e.g., R^2 above the range of ~0.8) between the actual and forecasted streamflow remains challenging [21]. To overcome this difficulty, generally, novel or hybrid ensemble models are being developed in combination with traditional non-linear streamflow forecasting models. Further, newly developed advanced metaheuristic techniques which are being used in other time-series forecasting applications (e.g., finance) from the machine learning domain are also occasionally being used in streamflow forecasting to determine if they can improve the prediction accuracy of streamflow forecast models [22,23]. Boosting algorithms are one such metaheuristic technique that has been recently applied to forecast streamflow [23–25].

Boosting algorithms like gradient tree boosting have been shown to forecast streamflow with high accuracy [23], and ensemble parallel tree boosting models like extreme gradient tree boosting (XGBoost) have also recently been used in related research [26,27]. Although very similar to RF in its structure and implementation, XGBoost varies by how the trees are developed and combined [28]. While RF is implemented by bagging where the final forecast is the average of all decision trees, XGBoost uses the error residuals from previous decision tree models to fit the subsequent models, and the final forecast is a weighted sum of all the tree forecasts. Similar to these models, a new model has been developed using gradient boosting on decision trees with categorical feature support, also referred to as CatBoost [29]. CatBoost has several advantages over traditional models, i.e., fast computing efficiency, native ability to handle categorical features, and the use of symmetric trees for swifter execution and to overcome overfitting by implementing ordered boosting [30]. Although CatBoost has not been applied in streamflow forecasting (to the best of the authors' knowledge), recent studies have documented the superior ability of CatBoost models in forecasting other hydro-climatic variables including weather and evapotranspiration [31,32].

In this study, three tree-based metaheuristic regression approaches, namely XGBoost, RF, and CatBoost, were used to model streamflow at three different locations in the Black Sea region in Turkey. The performance of all the three models was analyzed using different performance and error indices, and the results of each of these models were compared with ANN and nonlinear regression (NLR). Additionally, satellite rainfall data derived from Tropical Rainfall Measuring Mission (TRMM) were also used as model inputs to study if providing additional hydro-climatic variables improved the accuracy of the models. This paper presents some important observations regarding the use of the XGBoost and CatBoost models in streamflow forecasting and discusses how the performance of these models could be improved further in streamflow forecasting applications.

2. Methods and Materials

Three metaheuristic regression approaches, i.e., CB, RF, XGB, along with ANN and NLR were implemented in the present study to predict monthly streamflow considering precipitation data from TRMM. The modeling procedure is illustrated in Figure 1. The CB, RF and XGB methods are briefly explained in the following sections.

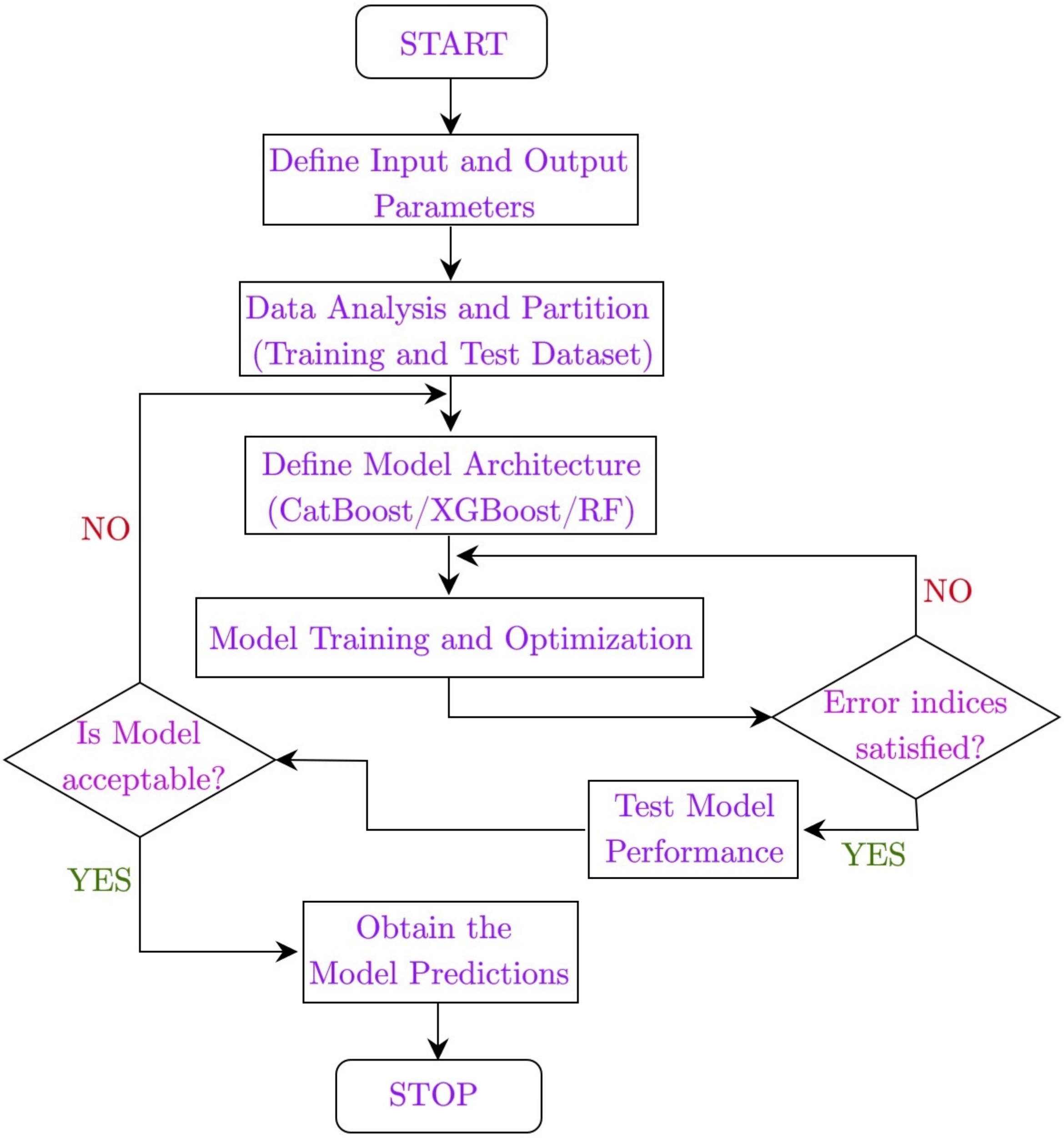

START
Define Input and Output Parameters
Data Analysis and Partition (Training and Test Dataset)
Define Model Architecture (CatBoost/XGBoost/RF)
Model Training and Optimization
Error indices satisfied?
NO
YES
Test Model Performance
Is Model acceptable?
NO
YES
Obtain the Model Predictions
STOP

Figure 1. The modeling procedure of the metaheuristic regression methods implemented in this study.

2.1. CatBoost

CatBoost a recent open-source boosting (ensemble strategy) algorithm proposed by Yandex engineers (documentation of CatBoost model is available at https://catboost.ai/) (accessed on 30 October 2022) [30–33]. It stems from the concepts of decision trees and gradient boosting. With oblivious ('Oblivious trees' are grown symmetrically, using the same features for splitting and learning criterion across each level of the tree.). decision trees as base predictors, CatBoost is well-balanced, less prone to overfitting, and saves significant time during testing phase. Let us consider dataset $Z = (a_i.b_i)_{i=1,\ldots,n}$, where $a_i = \left(a_i^1 \ldots\ldots.\, a_i^m\right)$ is a random vector of m features and $b_i \in \mathbb{R}$ is an output feature of either numerical or binary response. The data $(a_i.b_i)$ are independent and follow some unknown $N(\cdot,\cdot)$ distribution. A train function $F : \mathbb{R}^m \to \mathbb{R}$ that minimizes the expected loss $\mathcal{L}(F) = \mathbb{E}L(b.F(a))$ is the ultimate objective of any learning model, where L denotes a smooth loss function [29,30]. A sequence of relatively closer approximations $F^t : \mathbb{R}^m \to \mathbb{R}$, t = 0,1,2 . . . . is built iteratively in a greedy fashion using a gradient boosting procedure. Based on a generalized additive approach, F^t is derived from antecedent approximation F^{t-1}, such that $F^t = F^{t-1} + \alpha h^t$, where h^t is a base predictor function $(h^t : \mathbb{R}^m \to \mathbb{R})$ and α denotes the step size [30]. With the objective of minimizing the expected loss, the base predictor is usually opted from family of functions H:

$$h^t = \underset{h \in H}{\operatorname{argmin}} \mathcal{L}\left(F^{t-1} + h\right) = \underset{h \in H}{\operatorname{argmin}} \mathbb{E}\mathcal{L}\left(y.F^{t-1}(x) + h(x)\right) \tag{1}$$

The minimization problem is usually solved by functional gradient descent, either by considering a (negative) gradient step or by using the Newton second-order approximation method [29,33]. Generally, least-squares approximation is used to solve for $h^t(x)$. However, in CatBoost (an implementation of gradient boosting), decision tree 'h' is obtained as:

$$h^t h(x) = \sum_{k=1}^{K} m_k \mathbb{I}_{\{x \in D_k\}} \tag{2}$$

$h(x) = \sum_{k=1}^{K} m_k \mathbb{I}_{\{x \in D_k\}}$, where D_k are the disjointed regions that correspond to the tree's leaves and mk denotes the leaf values of the obtained trees [33].

CatBoost makes improvements to the gradient boosting procedure and employs a more effective 'Ordered Target Statistics' strategy to learn the model, making use of all training data. Hence, CatBoost outperforms even for heterogeneous data situations. For further in-depth details and mathematical concepts of CatBoost, readers may refer to the following literature [29,30,33,34].

2.2. eXtreme Gradient Boosting (XGBoost)

One of the fundamental issues in tree learning is to discover the best split; hence, eXtreme Gradient Boosting (XGBoost) (an extension to gradient boosted decision trees) was created, achieving superior results [35]. XGBoost employs a greedy algorithm that initiates from a single leaf and iteratively augments branches to the tree to find the best split [36]. It is not possible to train several trees in parallel using XGboost, but it can generate distinct tree nodes in parallel. The distributed weighted quantile sketch algorithm included in XGBoost aids in determining the best split points and handle weighted datasets. The weights of individual tree can be scaled down by a constant, thus reducing the impact of a single tree on the final score. To penalize the highly complex model, XGBoost employs both Lasso and Ridge Regression regularization. Additionally, it comes with a built-in cross-validation method at each iteration to avoid over-fitting. XGBoost has the advantages of effective tree pruning, parallel processing, and regularization. Finally, features like shrinkage and column subsampling speed up the computations of the parallel algorithm. For further in-depth details and mathematical concepts of XGBoost, the reader may refer to the following literature [35,37,38]. Figure 2 illustrates the structure of the XGBoost model.

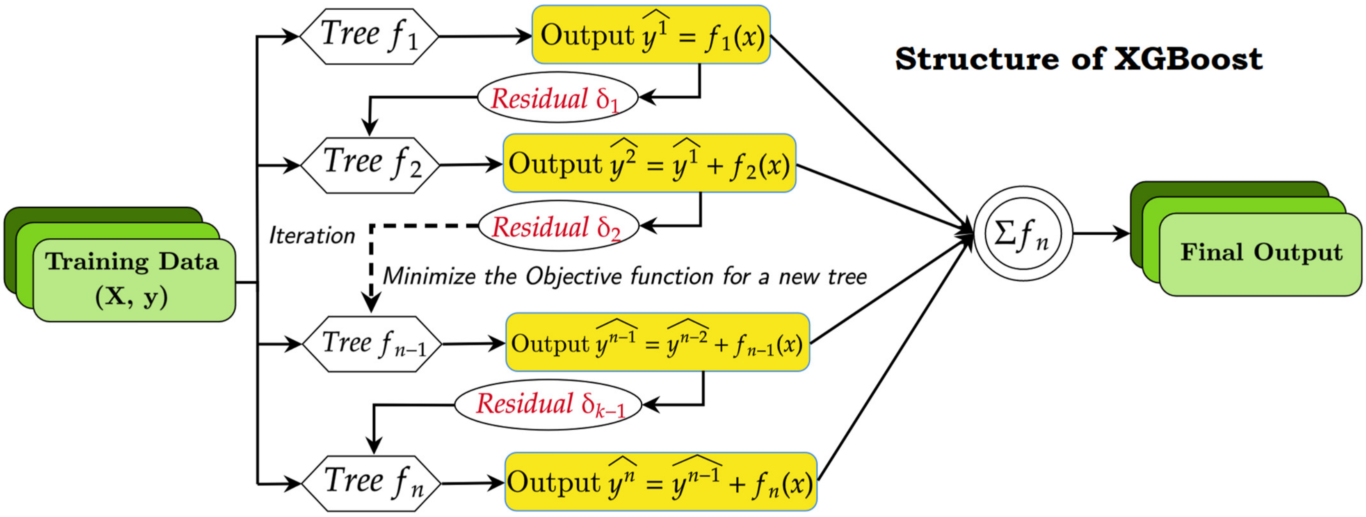

Structure of XGBoost
Training Data (X, y)
Tree f_1
Output $\widehat{y^1} = f_1(x)$
Residual δ_1
Tree f_2
Output $\widehat{y^2} = \widehat{y^1} + f_2(x)$
Residual δ_2
Iteration
Minimize the Objective function for a new tree
Tree f_{n-1}
Output $\widehat{y^{n-1}} = \widehat{y^{n-2}} + f_{n-1}(x)$
Residual δ_{k-1}
Tree f_n
Output $\widehat{y^n} = \widehat{y^{n-1}} + f_n(x)$
Σf_n
Final Output

Figure 2. Structure of the XGBoost model.

2.3. Random Forest (RF)

Random forest, an extension of the bagging method, is the most versatile supervised machine learning algorithm, wherein multiple individual decision trees are merged to form an ensemble [39]. The bagging approach selects a random sample of data from a training batch to generate several data samples and then train them independently. From 'k' number of data records, RF picks up 'n' number of data records to construct individual decision trees for each sample. One-third of the training sample is set aside as test data, referred to as the out-of-bag (oob) sample. Each decision tree provides an output and, based on majority voting or averaging, RF generates output for classification or regression tasks, respectively. Random forest allows for evaluating the importance of variables or their contribution to a model. When a variable is removed from a model, indices such as Gini importance and mean decrease in impurity (MDI) are commonly used to determine how much the model's accuracy has dropped. The RF model, unlike Decision Trees, is more robust to training sample selection and noise. Since it takes the average of all approximations from individual trees, overfitting is not seen due to the cancelling out of biases. For further in-depth details, the reader may refer to the following literature [39–42].

2.4. Case Study

The present study uses monthly mean streamflow data from three stations, Durucasu (station no: 1413, latitude: 36.11 N, longitude: 40.74 E, altitude: 301 m), Sutluce (station no: 1414, latitude: 36.12 N, longitude: 40.43 E, altitude: 510 m) and Kale (station no: 1402, latitude: 36.51 N, longitude: 40.77 E, altitude: 190 m), situated in Black Sea Region (BSR) of Turkey (Figure 3). The utilized data comprised continuous values throughout the period of 1998–2007; there were no gaps in the data from any of the stations. The highest rainfall in Turkey is observed in this region (BSR). The eastern part of the BSR receives 2200 mm of annual rainfall. This region has a wet-humid climate with a yearly average relative humidity of 71% and average temperatures of 4 °C and 22 °C in winter and summer, respectively. Yearly average total rainfall is about 842 mm and most of this (19.4%) occurs in summer [43]. Streamflow data were obtained from Turkish State Water Works. Precipitation data were obtained from the Tropical Rainfall Measuring Mission (TRMM), which provides continues satellite data over the BSR region. Such data were previously tested by comparing land data, and a high level of accuracy was observed by the researchers [43–47]. Table 1 sums up the statistical characteristics of the streamflow data. As shown, the streamflow data have high skewness, ranging from 1.60 to 2.43. The ranges of the training data do not cover those of the testing data; this could cause difficulties in predicting streamflow beyond the extreme values provided to the model in the training stage. The TRMM provides monthly precipitation data on grid bases. We selected the closest grids to the streamflow stations. Thus, for the Durucasu Station, TRMM from grid points #1524 (latitude: 36.125 N, longitude: 40.625 E) and #1602 (latitude: 36.125 N, longitude: 40.875 E) was used. For the Sutluce Station, data from grid #1446 (latitude: 36.125 N, longitude: 40.375 E), and for the Kale Station, data from grid #1525 (latitude: 36.375 N, longitude: 40.625 E) and grid #1526 (latitude: 36.625 N, longitude: 40.625 E) were utilized. The TRMM was launched in 1997. It is a joint project developed by NASA and JAXA (the space agency of Japan). It uses both active and passive microwave instruments with a low inclination orbit (35°). Therefore, TRMM is the foremost satellite in the world for the study of precipitation, storms and climate processes in the tropics (https://gpm.nasa.gov/sites/default/files/document_files/TRMMSenRevProp_v1.2.pdf) (accessed on 30 October 2022).

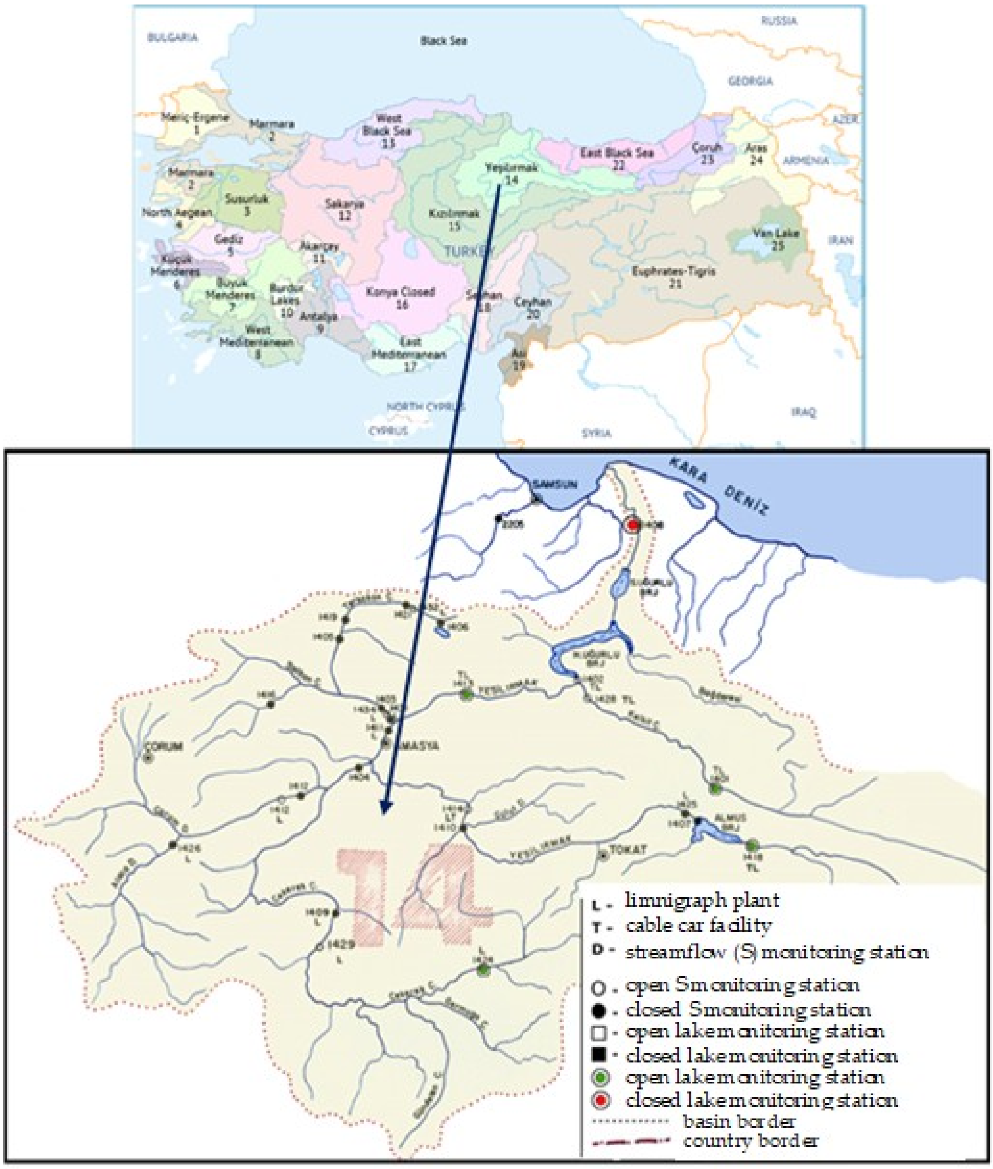


Figure 3. The location of the Durucasu (1413), Sutluce (1414) and Kale (1402), Yesilirmak Basin stations, situated in the Black Sea Region.

Table 1. Statistical properties of the streamflow data.

Station	Station No	Phase	Streamflow Data					
			Q_{max}	Q_{min}	Q_{mean}	Sk	CV	STD
Durucasu	1413	Test	173	9	40.1	2.11	0.96	38.3
		Train	169	4.6	32.5	2.43	0.9	28.6
Sutluce	1414	Test	39.6	5.9	14.1	1.54	0.56	7.7
		Train	37.5	4.5	12.2	1.87	0.53	6.42
Kale	1402	Test	334	43.5	104.2	2.38	0.65	68.6
		Train	387	47.7	126.7	1.6	0.60	76.4

Notes: Q_{max}: Maximum streamflow, Q_{min}: Minimum streamflow, Q_{mean}: Mean streamflow, Sk: Skewness, CV: Variation coefficient, STD: Standard deviation.

2.5. Application and Evaluation of the Methods

Three metaheuristic regression approaches, i.e., CB, RF and XGB, were compared in predictions of monthly streamflow, considering precipitation data obtained from TRMM. First, lagged streamflow data from three stations, Durucasu, Sutluce and Kale, were used as inputs to the models. Then, a periodicity component, indicated by the month number (MN) of the output (streamflow, Q at time t, Qt), was included in the input combinations. Finally, the precipitation acquired from TRMM was added into the inputs to explore its impact on accuracy of the models. In order to assess the performance of the implemented methods, the following statistics were employed:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_1^N (Q_o - Q_p)^2}{N}} \tag{3}$$

$$\mathrm{rRMSE} = 100\left(\frac{\mathrm{RMSE}}{\overline{Q_o}}\right) \tag{4}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_1^N \left|\left(Q_o - Q_p\right)\right| \tag{5}$$

$$E_{L,M} = 1 - \frac{\sum_1^N \left|Q_o - Q_p\right|}{\sum_1^N \left|Q_o - \overline{Q_o}\right|}, E_{L,M} \le 1 \tag{6}$$

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^N \left|Q_o - Q_p\right| \tag{7}$$

where RMSE is Root Mean Square Error, rRMSE is the relative RMSE, MAE is Mean Absolute Error, $E_{L,M}$ is the Legate and McCabe's Index, MAPE is the mean or average of the absolute percentage errors [47], N is the quantity of datasets, Q_o and Q_p are observed and predicted streamflow and $\overline{Q_o}$ denotes the observed mean value.

3. Application and Results

3.1. Predicting Monthly Streamflow of Durucasu Station

In Table 2, the accuracies of the metaheuristic regression methods are compared in predicting the monthly streamflow at the Durucasu Station for the test stage. In the table, S_1 in parenthesis is Scenario 1 involving Q_{t-1}, while S_{123} indicates the scenario of Q_{t-1}, Q_{t-2}, Q_{t-3}. The corresponding scenarios with periodicity are shown by S_1M or S_{123}M, respectively. The periodicity component is the month number of the streamflow output, varying from 1 (January) to 12 (December). Therefore, we used the abbreviation M for the scenarios involving periodicity information, while S_1P refers the S_1 with TRMM precipitation data. A comparison of the methods without considering TRMM data revealed that periodicity input improves the model accuracy in monthly streamflow prediction. For example, the improvement in the RMSE of the CB, RF and XGB for the first scenario (S_1) was by 14.2%, 4%, 7% and 26.4%, respectively. The percentages were calculated using relative error (RE) (RE = (Value 1 − Value 2)∗100/Value 1). The CB and XGB almost had the same accuracy and they performed than RF with respect to all evaluation statistics. The right part of the Table 2 clearly shows that considering TRMM precipitation considerably improved the efficiency of the implemented methods in predicting monthly streamflow. Adding precipitation input increased the accuracy of CB with S_1 input by 24%, 23%, 29% and 45% compared to RMSE, rRMSE, MAE and $E_{L,M}$, respectively. This improved the corresponding statistics by 35%, 36%, 34% and 49% for the RF(S_1) and 20%, 21%, 20% and 36% for the XGB(S_1) models, respectively. Similar to the discharge-based models, here, the periodicity also considerably improved the model efficiency. For example, improvements in RMSE, rRMSE, MAE and $E_{L,M}$ of 29, 28, 25 and 17% were observed for the CB with inputs of S_1 and TRMM precipitation, of 3.7%, 2.7%, 12.2% and 8.2% for the RF with the same inputs, and of 39, 38, 38 and 39% for the XGB with the same inputs. Among the metaheuristic regression models, XGB with S_1, periodicity and TRMM precipitation as inputs had the best accuracy, with the lowest RMSE (12.33 m^3/s), rRMSE (0.31) and MAE (8.77 m^3/s) and the highest $E_{L,M}$ (0.68) in predicting monthly streamflow. It was followed by the CB model with the same inputs. The equation of the NLR is:

$$Q_{1413} = 1.34\left(MN^{-0.33} + P^{1.17} + Q_{t-1}^{0.554}\right) \tag{8}$$

where Q_{1413} is the current streamflow at the Durucasu Station (Code: 1413), MN is the Month number, P is the TRMM Precipitation current month and $Q_{t-1}^{0.554}$ is the streamflow of one month prior.

Table 2. The accuracies of the CB, RF, XGB, ANN and NLR methods in predictions of monthly streamflow at the Durucasu Station (Code: 1413) in the testing phase.

Without TRMM Data							With TRMM Data						
Model (Scenario)	Model Inputs	RMSE	rRMSE	MAE	$E_{L,M}$	MAPE	Model (Scenario)	Model Inputs	RMSE	rRMSE	MAE	$E_{L,M}$	MAPE
CB (S_1)	Q_{t-1}	24.11	0.60	16.48	0.40	48.12	CB (S_1P)	Q_{t-1}, P	18.32	0.46	11.71	0.58	45.35
CB (S_{12})	Q_{t-1}, Q_{t-2}	26.08	0.65	15.90	0.42	47.5	CB (S_{12}P)	Q_{t-1}, Q_{t-2}, P	24.24	0.60	13.55	0.51	46.15
CB (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	26.76	0.67	17.18	0.38	48.19	CB (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	27.63	0.69	15.59	0.43	46.12
RF (S_1)	Q_{t-1}	23.19	0.58	16.21	0.41	50.2	RF (S_1P)	Q_{t-1}, P	15.03	0.37	10.63	0.61	46.15
RF (S_{12})	Q_{t-1}, Q_{t-2}	26.99	0.67	17.74	0.36	53.2	RF (S_{12}P)	Q_{t-1}, Q_{t-2}, P	17.60	0.44	11.43	0.59	48.26
RF (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	24.85	0.62	15.78	0.43	52.1	RF (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	17.41	0.43	11.64	0.58	47.15
XGB (S_1)	Q_{t-1}	25.33	0.63	17.60	0.36	46.12	XGB (S_1P)	Q_{t-1}, P	20.16	0.50	14.13	0.49	45.19
XGB (S_{12})	Q_{t-1}, Q_{t-2}	26.02	0.65	19.73	0.28	48.2	XGB (S_{12}P)	Q_{t-1}, Q_{t-2}, P	15.73	0.39	10.28	0.63	46.15
XGB (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	24.97	0.62	17.46	0.37	49.78	XGB (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	16.35	0.41	11.64	0.58	48.12
ANN (S_1)	Q_{t-1}	27.86	0.69	19.18	0.39	53.2	ANN (S_1P)	Q_{t-1}, P	22.18	0.55	15.40	0.53	50.23
ANN (S_{12})	Q_{t-1}, Q_{t-2}	28.10	0.70	21.31	0.31	55.45	ANN (S_{12}P)	Q_{t-1}, Q_{t-2}, P	16.99	0.42	11.31	0.68	52.13
ANN (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	26.72	0.66	18.86	0.40	57.8	ANN (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	17.49	0.44	12.45	0.64	52.14
NLR (S_1)	Q_{t-1}	30.65	0.76	20.71	0.42	55.18	NLR (S_1P)	Q_{t-1}, P	24.40	0.61	16.32	0.58	53.14

Table 2. *Cont.*

Without TRMM Data							With TRMM Data						
Model (Scenario)	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**	**Model (Scenario)**	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**
NLR (S_{12})	Q_{t-1}, Q_{t-2}	30.35	0.75	23.44	0.34	58.49	NLR (S_{12}P)	Q_{t-1}, Q_{t-2}, P	18.35	0.46	12.44	0.71	55.12
NLR (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	28.59	0.71	20.37	0.43	56.4	NLR (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	18.71	0.47	13.70	0.69	52.15
CB (S_1M)	Q_{t-1}, MN	20.69	0.52	13.74	0.50	42.2	CB (S_1MP)	Q_{t-1}, MN, P	13.05	0.33	8.79	0.68	25
CB (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	23.16	0.58	14.26	0.48	45.35	CB (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	24.04	0.60	13.11	0.51	24.5
CB (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	22.04	0.55	13.20	0.52	45.21	CB (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	25.11	0.63	13.83	0.50	25.6
RF (S_1M)	Q_{t-1}, MN	22.09	0.55	13.81	0.50	43.24	RF (S_1MP)	Q_{t-1}, MN, P	14.48	0.36	9.33	0.66	25.23
RF (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	21.95	0.55	13.36	0.52	48.26	RF (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	15.63	0.39	9.87	0.64	25.48
RF (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	22.35	0.56	13.02	0.53	48.97	RF (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	16.16	0.40	10.42	0.62	26.5
XGB (S_1M)	Q_{t-1}, MN	18.65	0.51	14.70	0.47	40.23	XGB (S_1MP)	Q_{t-1}, MN, P	12.33	0.31	8.77	0.68	29.12
XGB (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	18.07	0.45	12.69	0.54	41.24	XGB (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	14.26	0.36	9.26	0.66	28.45
XGB (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	18.1	0.45	11.54	0.58	43.5	XGB (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	14.02	0.35	10.22	0.63	28.75
ANN (S_1M)	Q_{t-1}, MN	21.28	0.53	14.29	0.48	42.12	ANN (S_1MP)	Q_{t-1}, MN, P	16.01	0.15	22.21	0.47	31.12
ANN (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	22.34	0.52	13.58	0.48	43.15	ANN (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	16.33	0.15	23.10	0.47	31.15
ANN (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	21.92	0.51	14.29	0.47	44.12	ANN (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	15.53	0.14	21.54	0.46	32.12
NLR (S_1M)	Q_{t-1}, MN	24.40	0.61	15.14	0.45	47.35	NLR (S_1MP)	Q_{t-1}, MN, P	30.26	0.29	23.38	0.45	32.14
NLR (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	23.67	0.58	15.29	0.46	48.2	NLR (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	29.05	0.29	22.44	0.43	32.17
NLR (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	24.64	0.61	14.38	0.45	47.65	NLR (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	29.96	0.29	23.85	0.47	35.14

Figure 4a compares the prediction accuracies of the metaheuristic regression models together with ANN and NLR in predictions of monthly streamflow at the Durucasu Station using the Taylor diagram. From such a diagram, we can compare the standard deviation (STD), RMSE and correlation of the model predictions. The XGB model with inputs of S_1, periodicity and TRMM precipitation had a closer STD to the measured one than to other methods, although this model was closely followed by CB with the same inputs.

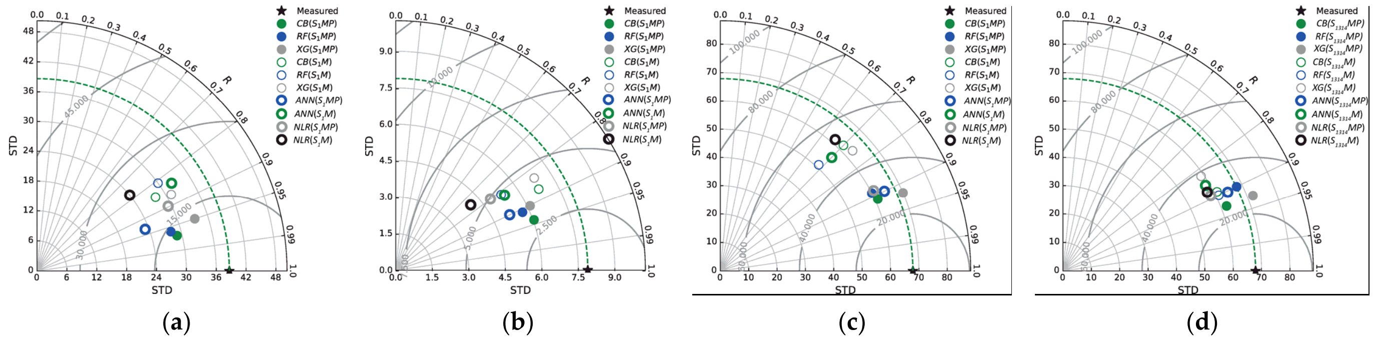


Figure 4. Taylor diagrams for the testing phase: (**a**) Durucasu Station, (**b**) Sutluce Station, (**c**,**d**) Kale Station.

Figure 5a,b compares the performance of three metaheuristic regression models with/without TRMM data for the first input scenario and involving periodicity (S_1M).

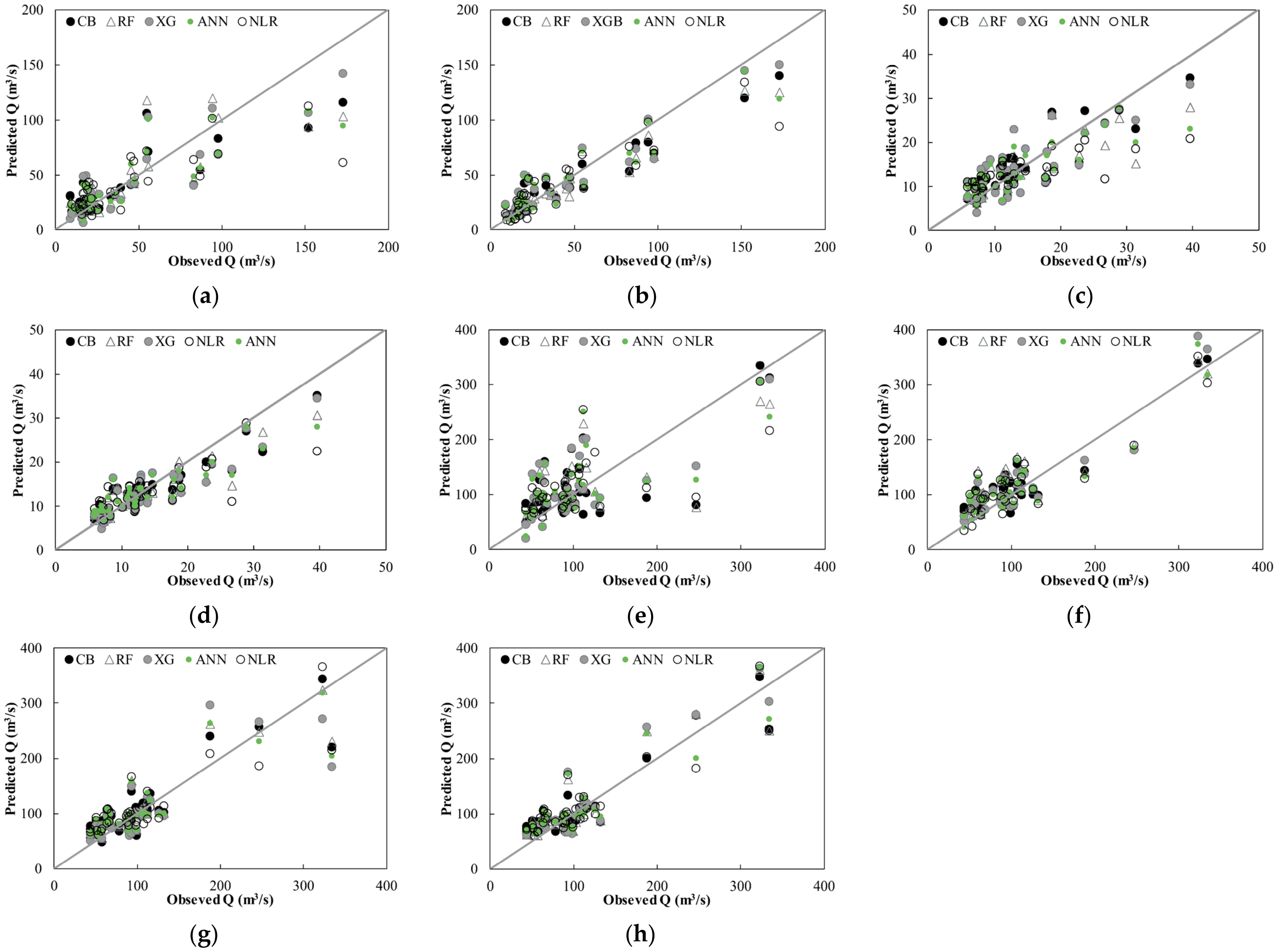
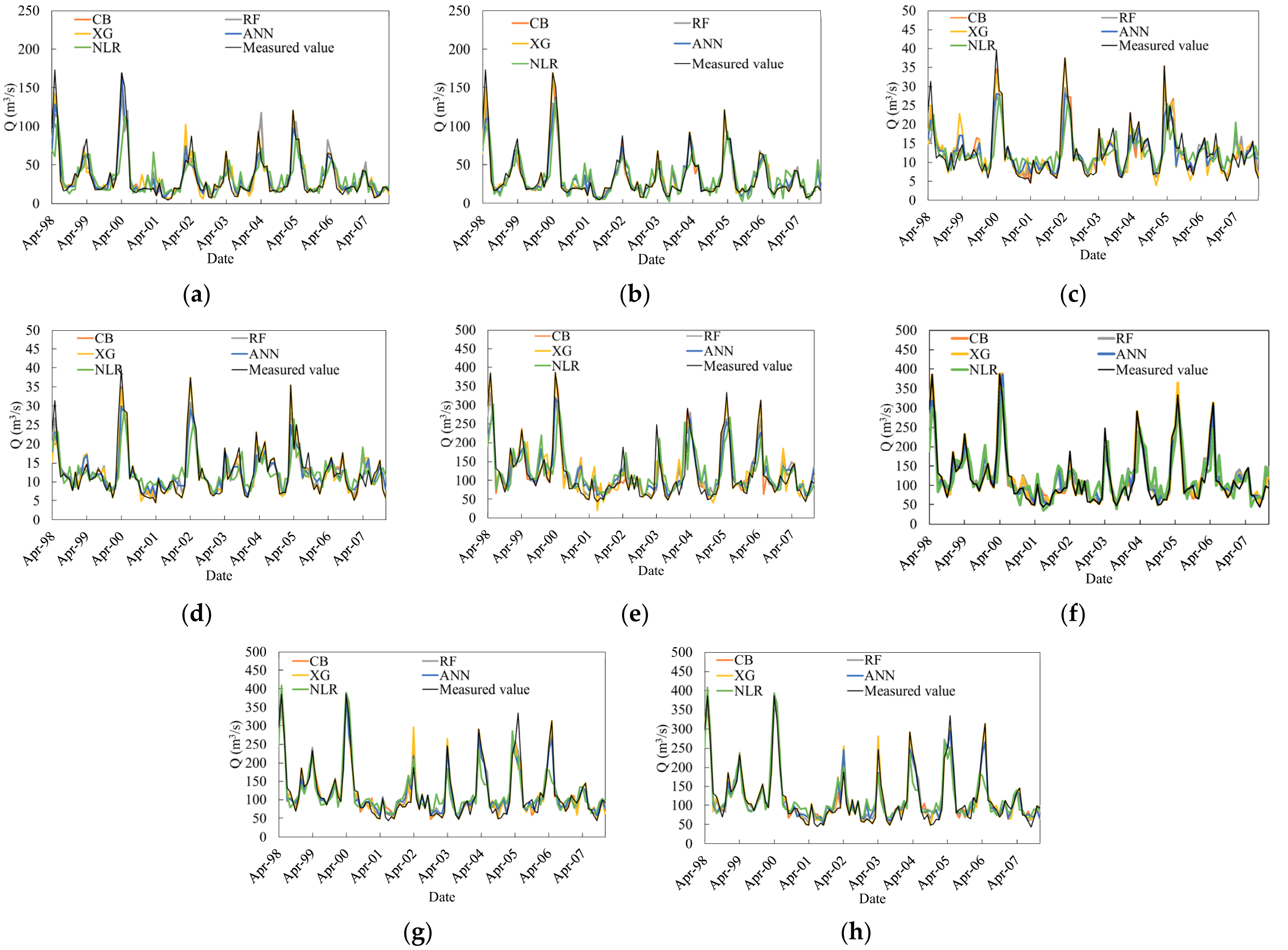


Figure 5. Scatter plots for (**a**) Durucasu (S_1M), (**b**) Durucasu (S_1MP) (**c**) Sutluce (S_1M), (**d**) Sutluce (S_1MP), (**e**) Kale (S_1M), (**f**) Kale (S_1MP), (**g**) Kale ($S_{1413}M$) and (**h**) Kale ($S_{1413}MP$).

From the two scatterplots, we can see that the use of TRMM data as input considerably improves the efficiency of all three methods, yielding less scattered predictions. Both the XGB model with inputs of S_1 and periodicity and the XGB model with inputs of S_1, periodicity and TRMM precipitation function better than the other models in predicting monthly the streamflow at the Durucasu Station. Figure 6a,b, compares the time variation of the model predictions in two cases (with/without TRMM data). As shown, getting precipitation information from TRMM data improves model performance in all ranges (low, mean and maximum flows).



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

(**g**) (**h**)

Figure 6. Time variation graphs of the observed and predicted stream flows by CB, RF, XG, ANN and NLR: (**a**) Durucasu (S_1M), (**b**) Durucasu (S_1MP) (**c**) Sutluce (S_1M), (**d**) Sutluce (S_1MP), (**e**) Kale (S_1M), (**f**) Kale (S_1MP), (**g**) Kale (S_{1314}M) and (**h**) Kale (S_{1314}MP).

3.2. Predicting Monthly Streamflow at Sutluce Station

Table 3 sums up the test results of the three metaheuristic regression methods in predicting monthly streamflow at the Sutluce Station. The left part of the table (without TRMM data) reveals that involving periodicity in the model inputs considerably improves performance; for example, the improvement in RMSE of the CB, RF and XGB models with S_1 input was by 52, 20 and 44%, respectively. The CB model with periodicity and two lagged streamflow data as inputs (Q_{t-1}, Q_{t-2}, MN) performed better than the other models, with the lowest RMSE (3.37 m^3/s), rRMSE (0.24), and MAE (2.58 m^3/s) and the highest $E_{L,M}$ (0.56) in the test stage. From the right part of Table 3, it can be observed that including TRMM precipitation data improved the model efficiency in both cases, with and without periodicity. Importing P data to the model input improved the RMSE, rRMSE, MAE and $E_{L,M}$ by 56, 56, 40 and 68% for the CB model with S_1 input, by 29, 28, 27 and 30% for the RF model with the same input and by 13.5, 14.3, 4.9 and 25% for the XGB model with the same input, respectively. For the CB model, having periodicity, TRMM precipitation and one lagged streamflow data as inputs (Q_{t-1}, MN, P) offered better performance than the other models. Similar to the previous station, here, an improvement was seen when periodicity was used as input; increases in RMSE, rRMSE, MAE and $E_{L,M}$ were by 38, 39, 42 and 70% for the CB model with inputs of S_1 and TRMM precipitation, by 20, 19, 18 and 3719% for the RF model with the same input and by 47, 47, 44 and 231% for the XGB model with the same input, respectively. The equation of the NLR is:

$$Q_{1414} = 0.98\left(MN^{-0.28} + P^{0.096} + Q_{t-1}^{0.84}\right) \tag{9}$$

where Q_{1414} is the current streamflow at the Sutluce Station (Code: 1414).

Table 3. The accuracies of the CB, RF, XGB, ANN and NLR methods in predictions of monthly streamflow at the Sutluce Station (Code: 1414) in the testing phase.

Without TRMM Data							With TRMM Data						
Model (Scenario)	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**	**Model (Scenario)**	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**
CB (S_1)	Q_{t-1}	7.83	0.56	5.18	0.12	53.2	CB (S_1P)	Q_{t-1}, P	5.02	0.36	3.71	0.37	45.2
CB (S_{12})	Q_{t-1}, Q_{t-2}	4.91	0.35	3.46	0.41	54.8	CB (S_{12}P)	Q_{t-1}, Q_{t-2}, P	5.14	0.37	3.41	0.42	46.5
CB (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	4.96	0.35	3.41	0.42	55.4	CB (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	5.59	0.4	3.56	0.40	46.8
RF (S_1)	Q_{t-1}	5.77	0.41	4.01	0.32	55.2	RF (S_1P)	Q_{t-1}, P	4.47	0.32	3.16	0.46	48.5
RF (S_{12})	Q_{t-1}, Q_{t-2}	4.98	0.35	3.35	0.43	55.6	RF (S_{12}P)	Q_{t-1}, Q_{t-2}, P	4.20	0.3	3.05	0.48	48.9
RF (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	5.17	0.37	3.63	0.38	56	RF (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	4.69	0.33	3.32	0.44	47.5
XGB (S_1)	Q_{t-1}	7.84	0.56	5.18	0.12	52.32	XGB (S_1P)	Q_{t-1}, P	6.91	0.49	4.94	0.16	43.5
XGB (S_{12})	Q_{t-1}, Q_{t-2}	6.12	0.43	4.46	0.24	53.6	XGB (S_{12}P)	Q_{t-1}, Q_{t-2}, P	4.88	0.35	3.40	0.42	42.9
XGB (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	5.30	0.38	3.88	0.34	53.87	XGB (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	5.39	0.38	3.61	0.39	44.5
ANN (S_1)	Q_{t-1}	8.62	0.62	5.59	0.13	57.2	ANN (S_1P)	Q_{t-1}, P	7.60	0.54	5.38	0.18	52.3
ANN (S_{12})	Q_{t-1}, Q_{t-2}	6.61	0.47	4.77	0.26	58.9	ANN (S_{12}P)	Q_{t-1}, Q_{t-2}, P	5.27	0.38	3.67	0.46	53.2
ANN (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	5.67	0.41	4.27	0.36	57.4	ANN (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	5.77	0.41	3.86	0.43	54.1
NLR(S_1)	Q_{t-1}	9.48	0.68	5.98	0.14	57.6	NLR (S_1P)	Q_{t-1}, P	8.36	0.60	5.92	0.19	51.6
NLR (S_{12})	Q_{t-1}, Q_{t-2}	7.14	0.51	5.20	0.28	58.2	NLR (S_{12}P)	Q_{t-1}, Q_{t-2}, P	5.69	0.41	4.00	0.49	53.1
NLR (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	6.07	0.43	4.65	0.39	58.4	NLR (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	6.17	0.44	4.09	0.46	52.6
CB (S_1M)	Q_{t-1}, MN	3.80	0.27	2.91	0.51	42.15	CB (S_1MP)	Q_{t-1}, MN, P	3.10	0.22	2.16	0.63	15.8
CB (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	3.37	0.24	2.58	0.56	43.15	CB (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	4.08	0.29	2.82	0.52	16.3
CB (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	4.49	0.32	2.91	0.51	42.18	CB (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	4.50	0.32	2.86	0.51	15.8
RF (S_1M)	Q_{t-1}, MN	4.63	0.33	3.15	0.51	43.32	RF (S_1MP)	Q_{t-1}, MN, P	3.60	0.26	2.58	0.63	18
RF (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	4.43	0.31	2.94	0.50	44.18	RF (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	3.78	0.27	2.59	0.56	19.5

Table 3. *Cont.*

Without TRMM Data							With TRMM Data						
Model (Scenario)	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**	**Model (Scenario)**	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**
RF (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	4.56	0.32	2.99	0.49	45.78	RF (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	3.88	0.28	2.63	0.55	18.9
XGB (S_1M)	Q_{t-1}, MN	4.37	0.31	3.59	0.51	40.59	XGB (S_1MP)	Q_{t-1}, MN, P	3.65	0.26	2.75	0.53	20.8
XGB (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	4.38	0.31	2.98	0.50	39.18	XGB (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	3.67	0.25	2.76	0.59	21.5
XGB (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	4.17	0.3	2.86	0.52	40.12	XGB (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	3.77	0.27	2.52	0.59	22.6
ANN (S_1M)	Q_{t-1}, MN	27.86	1.97	5.28	0.42	42	ANN (S_1MP)	Q_{t-1}, MN, P	3.92	0.28	1.03	0.54	21.1
ANN (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	27.30	1.95	5.28	0.42	43.5	ANN (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	3.80	0.27	1.08	0.54	21.6
ANN (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	26.75	1.91	5.49	0.43	42.26	ANN (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	4.12	0.29	1.05	0.53	23.5
NLR(S_1M)	Q_{t-1}, MN	5.41	0.38	3.59	0.39	43	NLR (S_1MP)	Q_{t-1}, MN, P	5.01	0.36	3.34	0.43	25.3
NLR (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	5.46	0.39	3.45	0.39	43.6	NLR (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	4.81	0.34	3.37	0.43	24.2
NLR (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	5.25	0.37	3.70	0.37	42.5	NLR (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	4.81	0.34	3.27	0.43	25.6

The Taylor diagram provided in Figure 4b shows that the CB model with inputs of S_1, periodicity and TRMM precipitation had lower RMSE, higher correlation and closer STD to the measured one than the other models. From the scatterplots in Figure 5c,d, it is clear that the CB with inputs of S_1, periodicity and TRMM precipitation had less scattered predictions. Additionally, it is clear that the use of TRMM data considerably improved the accuracy of the models. As shown from time variation graphs in Figure 6c,d, the models utilizing TRMM data could follow the measured streamflow much more closely than the discharge-based models.

3.3. *Predicting Monthly Streamflow at the Kale Station*

A comparison of metaheuristic regression methods in predicting monthly streamflow at the Kale Station is made in Table 4 for the test stage. It is apparent from the table (see the left part, without TRMM data) that considering periodicity in the input considerably improved the efficiency of the various methods; for example, improvements in RMSE, rRMSE, MAE and $E_{L,M}$ were by 31, 30, 32 and 158% for the CB model with S_1 input, by 26, 25 28 and 174% for the RF model with the same input and by 43, 43, 40 and 133% for the XGB model with the same input, respectively. Among the three metaheuristic regression methods, the XGB model with two lagged streamflow data and periodicity (Q_{t-1}, Q_{t-2}, MN) as inputs offered the best accuracy, with the lowest RMSE (44 m^3/s), rRMSE (0.42) and MAE (30.03 m^3/s) and the highest $E_{L,M}$ (0.29) in the test stage. It is apparent from the second part of the Table 4 (see the right part, with TRMM data) that the use of P data acquired from TRMM considerably improved the accuracy both with and without periodicity. For example, it improved the RMSE, rRMSE, MAE and $E_{L,M}$ by 29, 29, 100 and 6373% for the CB model with S_1 input, by 29, 28 100 and 8968% for the RF model with the same input and by 39, 40, 100 and 3977% for the XGB model with the same input, respectively. The equation of the NLR is:

$$Q_{1402} = 3.24\left(MN^{-0.06} + P^{1.19} + Q_{t-1}^{0.68}\right) \tag{10}$$

where Q_{1402} is the current streamflow at the Kale Station (Code: 1402).

Table 4. Accuracies of the CB, RF, XGB, ANN and NLR methods in predictions of monthly streamflow at the Kale Station (Code: 1402) in the testing phase.

Without TRMM Data							With TRMM Data						
Model (Scenario)	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	$E_{L,M}$	**MAPE**	**Model (Scenario)**	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	$E_{L,M}$	**MAPE**
CB (S_1)	Q_{t-1}	71.88	0.69	53.03	−0.26	42.3	CB (S_1P)	Q_{t-1}, P	51.41	0.49	36.7	−16.83	39.5
CB (S_{12})	Q_{t-1}, Q_{t-2}	61.36	0.59	44.83	−0.06	45.2	CB (S_{12}P)	Q_{t-1}, Q_{t-2}, P	50.26	0.48	37.09	0.12	38.6
CB (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	63.89	0.61	49.82	−0.18	44.8	CB (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	56.25	0.54	44.09	−0.05	37.6
RF (S_1)	Q_{t-1}	66.98	0.64	50.21	−0.19	45.3	RF (S_1P)	Q_{t-1}, P	47.51	0.46	34.15	−17.23	43.5
RF (S_{12})	Q_{t-1}, Q_{t-2}	59.93	0.57	44.52	−0.06	48.5	RF (S_{12}P)	Q_{t-1}, Q_{t-2}, P	51.04	0.49	37.66	0.11	42.5
RF (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	63.25	0.61	46.81	−0.11	46.8	RF (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	55.60	0.53	40.59	0.04	43.8
XGB (S_1)	Q_{t-1}	83.01	0.8	60.21	−0.43	43.4	XGB (S_1P)	Q_{t-1}, P	50.49	0.48	37.83	−17.53	40.5
XGB (S_{12})	Q_{t-1}, Q_{t-2}	66.86	0.64	50.37	−0.20	43.5	XGB (S_{12}P)	Q_{t-1}, Q_{t-2}, P	56.27	0.54	40.95	0.03	40.15
XGB (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	66.67	0.64	48.71	−0.16	44.8	XGB (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	58.15	0.56	42.81	−0.02	41.5
ANN (S_1)	Q_{t-1}	91.31	0.88	64.42	−0.46	48.9	ANN (S_1P)	Q_{t-1}, P	55.54	0.54	41.23	−19.11	43.5
ANN (S_{12})	Q_{t-1}, Q_{t-2}	72.21	0.70	55.41	−0.22	50.2	ANN (S_{12}P)	Q_{t-1}, Q_{t-2}, P	60.77	0.59	44.64	0.03	44.8
ANN (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	71.34	0.69	53.09	−0.18	49.5	ANN (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	62.22	0.60	46.66	−0.02	44.9
NLR (S_1)	Q_{t-1}	100.44	0.97	68.93	−0.50	50.2	NLR (S_1P)	Q_{t-1}, P	61.09	0.59	44.94	−20.83	44.32
NLR (S_{12})	Q_{t-1}, Q_{t-2}	77.99	0.75	60.40	−0.24	53.2	NLR (S_{12}P)	Q_{t-1}, Q_{t-2}, P	65.63	0.63	48.66	0.03	43.9
NLR (S_{123})	Q_{t-1}, Q_{t-2}, Q_{t-3}	76.33	0.74	56.81	−0.20	54.9	NLR (S_{123}P)	Q_{t-1}, Q_{t-2}, Q_{t-3}, P	66.58	0.64	50.39	−0.02	43.72
CB (S_1M)	Q_{t-1}, MN	49.68	0.48	35.86	0.15	38.1	CB (S_1MP)	Q_{t-1}, MN, P	29.11	0.28	24.87	0.41	30.6
CB (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	49.78	0.48	37.32	0.11	36.5	CB (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	44.30	0.43	33.19	0.21	31.2
CB (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	50.17	0.48	37.45	0.11	38.9	CB (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	44.84	0.43	36.02	0.15	30.8
RF (S_1M)	Q_{t-1}, MN	49.68	0.48	36.20	0.14	38	RF (S_1MP)	Q_{t-1}, MN, P	33.41	0.32	26.79	0.41	32.35
RF (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	53.77	0.52	40.00	0.05	40.3	RF (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	46.99	0.45	34.55	0.18	33.54
RF (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	58.40	0.56	42.03	0.00	39.5	RF (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	51.63	0.5	36.86	0.13	35.6
XGB (S_1M)	Q_{t-1}, MN	47.71	0.46	36.29	0.14	40.1	XGB (S_1MP)	Q_{t-1}, MN, P	29.32	0.28	23.80	0.41	28.5

Table 4. *Cont.*

Without TRMM Data							With TRMM Data						
Model (Scenario)	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**	**Model (Scenario)**	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**
XGB (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	44.00	0.42	30.03	0.29	40.3	XGB (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	45.25	0.43	31.59	0.25	28.4
XGB (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	45.36	0.44	33.94	0.19	42.8	XGB (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	42.30	0.41	30.22	0.28	26.5
ANN (S_1M)	Q_{t-1}, MN	33.58	0.32	36.20	0.14	43.4	ANN (S_1MP)	Q_{t-1}, MN, P	31.21	0.29	24.74	0.44	35.8
ANN (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	32.24	0.31	36.20	0.14	45.2	ANN (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	30.27	0.29	25.98	0.43	35.3
ANN (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	32.57	0.31	37.29	0.14	43.5	ANN (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	29.65	0.29	24.25	0.46	35.7
NLR (S_1M)	Q_{t-1}, MN	50.66	0.49	36.08	0.14	41.3	NLR (S_1MP)	Q_{t-1}, MN, P	32.13	0.30	26.51	0.37	37.2
NLR (S_{12}M)	Q_{t-1}, Q_{t-2}, MN	52.69	0.51	35.36	0.13	43.5	NLR (S_{12}MP)	Q_{t-1}, Q_{t-2}, MN, P	32.45	0.31	27.31	0.35	37.8
NLR (S_{123}M)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN	52.18	0.50	36.44	0.14	44.4	NLR (S_{123}MP)	Q_{t-1}, Q_{t-2}, Q_{t-3}, MN, P	33.42	0.32	25.18	0.36	38.9

Among the three metaheuristic regression methods utilizing TRMM data as inputs, the XGB(S_1MP) and CB(S_1MP) with one lagged streamflow, TRMM precipitation and periodicity (Q_{t-1}, P, MN) as inputs had almost the same accuracy, with both performing better than the RF model in the test stage. By using periodicity information, considerable improvements were observed; for example, improvements in the RMSE, rRMSE, MAE and $E_{L,M}$ were by 43, 43, 32 and 102% for the CB model with inputs of S_1 and TRMM precipitation by 30, 30, 21.55 and 102% for the RF model with the same input and by 42, 42, 37 and 102% for the XGB with the same input, respectively. By importing periodicity information, considerable improvements were observed; for example, improvements in the RMSE, rRMSE, MAE and $E_{L,M}$ were by 43, 43, 32 and 102% for the CB inputs of S_1 and TRMM precipitation, by 30, 30, 21.55 and 102% for the RF with the same input and by 42, 42, 37 and 102% for the XGB with the same input, respectively.

Figure 3 provides a Taylor diagram comparing the three methods with respect to correlation coefficient(R), RMSE and STD. The XGB model, with inputs of S_1, periodicity and TRMM precipitation had a closer STD to the measured one than the others, closely followed by the CB model with the same input. As observed from Figure 4c,d, the use of TRMM data considerably improved the accuracy of all models. We can also see this improvement in time variation graphs provided in Figure 5e,f.

3.4. Predicting Monthly Streamflow at the Kale Station Using Upstream Data

Predicting monthly streamflow using data from upstream stations is essential. In some cases, data are missing from some stations because of technical problems, especially in the developing countries like Turkey. In this section, three metaheuristic regression methods are employed to find an efficient prediction model. Monthly streamflow data from the Kale Station were predicted using data of two upstream stations, Durucasu and Sutluce. Here, periodicity and TRMM data were also considered. Table 5 compares the accuracy of three methods with respect to some evaluation criteria utilized in the previous applications. It is clear from the table that in both cases (with/without TRMM data), considering periodicity generally improved the accuracy; for example, the RMSE decreased from 33.51 m^3/s to 30.48 m^3/s for the CB(S_{1314}) model, from 32.60 m^3/s to 28.04 m^3/s for the RF(S_{1314}) model and from 48.59 m^3/s to 38.63 m^3/s for the XGB(S_{1314}) model. In both cases (with/without periodicity), adding TRMM data improved the efficiency; for example, a decrease was observed in RMSE from 30.48 m^3/s to 25.79 m^3/s for the CB(S_{1314}M) model, from 32.60 m^3/s to 27.78 m^3/s for the RF(S_{1314}) model and from 48.59 m^3/s to 28.22 m^3/s for the XGB(S_{1314}) model. Among the implemented models, CB(S_{1314}MP) produced the best streamflow predictions, with the lowest RMSE (25.79 m^3/s) rRMSE (0.25), MAE (20.39 m^3/s) and the highest $E_{L,M}$ (0.52). The equation of the NLR is:

$$Q_{1402} = 6.53\left(MN^{-0.85} + P^{0.3} + Q_{1414}^{0.004} + Q_{1413}^{0.92}\right) \qquad (11)$$

where Q_{1402} is the current streamflow at the Kale Station (Code: 1402), Q_{1414} is the current streamflow at Sutluce Station and Q_{1413} is the current streamflow at the Durucasu Station.

Table 5. Accuracies of the CB, RF, XGB, ANN and NLR methods in predictions of monthly streamflow at the Kale Station (Code:1402) using upstream data from the Durucasu (Code:1413) and Sutluce (Code: 1414) stations in the testing phase.

Without TRMM Data							With TRMM Data						
Model (Scenario)	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**	**Model (Scenario)**	**Model Inputs**	**RMSE**	**rRMSE**	**MAE**	**$E_{L,M}$**	**MAPE**
CB (S_{1314})	Q_{1413}, Q_{1414}	33.51	0.32	22.84	0.46	23.5	CB (S_{1314}P)	Q_{1413}, Q_{1414}, P	31	0.3	23.76	0.44	18.5
CB (S_{1314}M)	Q_{1413}, Q_{1414}, MN	30.48	0.29	22.72	0.46	21	CB (S_{1314}MP)	Q_{1413}, Q_{1414}, MN, P	25.8	0.25	20.39	0.52	15.3
RF (S_{1314})	Q_{1413}, Q_{1414}	32.60	0.31	21.30	0.49	26.8	RF (S_{1314}P)	Q_{1413}, Q_{1414}, P	27.7	0.27	20.44	0.52	23.1
RF (S_{1314}M)	Q_{1413}, Q_{1414}, MN	29.04	0.27	18.85	0.55	25	RF (S_{1314}MP)	Q_{1413}, Q_{1414}, MN, P	28.7	0.28	21.68	0.52	18.8
XGB (S_{1314})	Q_{1413}, Q_{1414}	48.59	0.47	30.38	0.28	23.5	XGB (S_{1314}P)	Q_{1413}, Q_{1414}, P	28.2	0.27	21.67	0.49	18.5
XGB (S_{1314}M)	Q_{1413}, Q_{1414}, MN	38.63	0.37	25.03	0.41	21	XGB (S_{1314}MP)	Q_{1413}, Q_{1414}, MN, P	27.1	0.26	21.22	0.52	15.65
ANN (S_{1314})	Q_{1413}, Q_{1414}	53.45	0.52	33.42	0.31	29.5	ANN (S_{1314}P)	Q_{1413}, Q_{1414}, P	31.02	0.30	23.62	0.52	18.5
ANN (S_{1314}M)	Q_{1413}, Q_{1414}, MN	41.72	0.40	27.53	0.45	27.5	ANN (S_{1314}MP)	Q_{1413}, Q_{1414}, MN, P	29.27	0.28	23.34	0.56	25.45
NLR (S_{1314})	Q_{1413}, Q_{1414}	58.80	0.57	36.09	0.34	33.2	NLR (S_{1314}P)	Q_{1413}, Q_{1414}, P	34.12	0.33	25.51	0.56	29.5
NLR (S_{1314}M)	Q_{1413}, Q_{1414}, MN	45.06	0.43	29.46	0.48	30.3	NLR (S_{1314}MP)	Q_{1413}, Q_{1414}, MN, P	31.61	0.30	24.97	0.61	27

The prediction results of the three metaheuristic regression methods are illustrated in Figure 4d in a Taylor diagram. As shown in the diagram, the CB(S_{1314}MP) model achieved better accuracy with closer STD to the measured one and lower RMSE and higher correlation than the other methods. In this regard, this model was closely followed by the XG(S_{1314}MP) model. From the scatter plots in Figure 5g,h and time variation graphs in Figure 6g,h, the performance improvement by using TRMM data with the implemented models is clearly seen.

4. Discussion

In the presented study, three metaheuristic regression methods, CB, RF and XGB, were implemented for monthly streamflow predictions. These approaches were then compared with ANN and NLR methods. The applicability of TRMM precipitation data as inputs to the aforementioned models was investigated by considering different input scenarios comprising lagged streamflow as inputs as well as periodicity information (month number). The overall results indicate that considering TRMM precipitation data as inputs to the metaheuristic regression methods considerably improved their accuracy in monthly streamflow predictions, i.e., improvements in RMSE and MAE of the CB models having one lagged streamflow as input were by 24 and 29% for the Durucasu Station, by 56 and 40% for the Sutluce Station and by 29 and 100% for the Kale Station. This implies that such data are very useful in complex monthly streamflow predictions, especially in developing countries, where precipitation measurements are not available or may be missing altogether for technical reasons. These results are in agreement with the literature [45–47]. The accuracy of TRMM precipitation data was assessed in [45]. The authors of that report compared the monthly TRMM precipitation data covering the period of 1998–2010 with rain gauges from 16 meteorological stations in the Yarlung Zangbo River Basin and reported that there was a strong correlation and little numerical biases between TRMM precipitation data and rain gauges. By comparing its precipitation data with the 149 rainfall stations in Tunisia for a 16-year period (1998–2013), the performance of TRMM was assessed [46]. The authors found strong correlation between them.

It is observed from the results that adding a periodicity component to the inputs of the models improved their prediction accuracy, both with and without TRMM data. Improvements in the RMSE and MAE of the CB(S_1) models were by 14 and 17% for the Durucasu Station, by 52 and 44% for the Sutluce Station and by 31 and 32% for the Kale Station. Similar observations were reported in a previous study [48]. Monthly streamflow of a mountainous basin using machine learning methods (e.g., MARS, GMDH) was predicted in a previous study [48]. The authors of that paper reported that the use of periodicity information in the model inputs generally improved the accuracy of their predictions.

The results of streamflow predictions using upstream data revealed that such data can provide more information than local data. A comparison of Tables 4 and 5 shows that the use of upstream data (from the Durucasu and Sutluce stations) without local station data (i.e., the Kale Station) considerably improved the model efficiency with respect to RMSE, rRMSE, MAE and $E_{L,M}$. Improvements in the RMSE and rRMSE of the CB model with one lagged streamflow as input and without TRMM data were 53 and 54%, while the corresponding percentages were 40 and 39% for the same model with TRMM data. These are very useful findings for predictions of monthly streamflow, especially in the basins where limited measurements are made.

From a comparison of the tables, it may be seen that the addition of more lagged streamflow as inputs to the implemented models reduced their accuracy. These results are in direct agreement with those of previous studies [49,50]. According to the reports provided by the abovementioned references, increasing input quantity does not guarantee better predictions and, in some cases, it may negatively affect variance. In other words, increasing input quantity may create a more complex model with poor prediction accuracy.

Three neural network methods, i.e., feed forward neural networks (FFNN), generalized regression neural networks (GRNN) and radial basis function (RBF) were used to predict monthly streamflow at two stations, Gerdelli and Isakoy, in Turkey [51]. The GRNN provided the best accuracy with the lowest RMSE of 9.25 and 14.2 m^3/s for the Gerdelli (Q_{mean} = 12.29 m^3/s) and Isakoy (Q_{mean} = 17.86 m^3/s) stations, respectively. Two neuro-fuzzy methods were applied for predictions of monthly streamflow at two additional stations, Besiri and Baykan, in Turkey; the best results were obtained from the subclustering-based neuro-fuzzy metho, which obtained an RMSE of 32.8 and 9.36 m^3/s for the Besiri (Q_{mean} = 51.82 m^3/s) and Baykan (Q_{mean} = 21.36 m^3/s) stations, respectively [52]. In the present study, the CB, as the best method, produced an RMSE of 12.33, 3.1 and

29.11 m^3/s for the Durucasu (Q_{mean} = 40.1 m^3/s), Sutluce (Q_{mean} = 14.1 m^3/s) and Kale (Q_{mean} = 104.2 m^3/s) stations. This proves the accuracy of the implemented metaheuristic regression method (CB) in monthly streamflow predictions. In addition, the CB method has a simpler structure than the FFNN, GRNN, RBF and neuro-fuzzy methods.

5. Conclusions

In this study, the viability of three metaheuristic regression methods was investigated for monthly streamflow predictions using streamflow data from three stations in Turkey and satellite precipitation data from TRMM. The results were also compared with those obtained using the ANN and NLR models. The outcomes revealed that satellite data are very useful for monthly streamflow predictions, considerably improving the ability of metaheuristic regression methods (e.g., CB, RF and XGB). Our assessment of the methods with respect to statistical measures (e.g., RMSE, rRMSE, MAE, EL, M and CA) and visual inspections (Taylor diagrams, scatterplots and hydrographs) revealed that the CB method generally performed better than the XGB, RF, ANN and NLR methods. Additional improvement was observed by introducing periodicity information to the models. This input, involving month number, is very easy to employ and its usage is highly recommended for engineers and scholars. Including TRMM precipitation as input considerably improved the accuracy of implemented methods. This satellite data is highly accurate and its usage in streamflow predictions is strongly recommended by the authors. Monthly streamflow at the downstream station was successfully predicted by the CB and XGB methods using upstream data. In this application, the use of TRMM precipitation information and a periodicity component provided additional accuracy to the implemented models. These findings may provide useful information for managers and decision makers, especially in developing countries, where precipitation data are missing or absent altogether because of technical issues.

Author Contributions: M.M., methodology, investigation, formal analysis; A.M., formal analysis, writing—review and editing; S.R.N., methodology, resources, writing—review and editing; C.K., writing—review and editing; O.K., methodology, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to thank the staff of the Turkish State Water Works who provided the streamflow data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Edwards, P.J.; Williard, K.W.; Schoonover, J.E. Fundamentals of watershed hydrology. *J. Contemp. Water Res. Educ.* **2015**, *154*, 3–20. [CrossRef]
2. Davie, T. *Fundamentals of Hydrology*, 2nd ed.; Routledge: London, UK, 2019.
3. Chegwidden, O.S.; Rupp, D.E.; Nijssen, B. Climate change alters flood magnitudes and mechanisms in climatically-diverse headwaters across the northwestern United States. *Environ. Res. Lett.* **2020**, *15*, 094048. [CrossRef]
4. Goeking, S.A.; Tarboton, D.G. Forests and water yield: A synthesis of disturbance effects on streamflow and snowpack in western coniferous forests. *J. For.* **2020**, *118*, 172–192. [CrossRef]
5. Naz, B.S.; Kao, S.C.; Ashfaq, M.; Gao, H.; Rastogi, D.; Gangrade, S. Effects of climate change on streamflow extremes and implications for reservoir inflow in the United States. *J. Hydrol.* **2019**, *556*, 359–370. [CrossRef]
6. Valenzuela-Aguayo, F.; McCracken, G.R.; Manosalva, A.; Habit, E.; Ruzzante, D.E. Human-induced habitat fragmentation effects on connectivity, diversity, and population persistence of an endemic fish, Percilia irwini, in the Biobío River basin (Chile). *Evol. Appl.* **2020**, *13*, 794–807. [CrossRef]

7. Allen, G.H.; Pavelsky, T.M. Global extent of rivers and streams. *Science* **2018**, *361*, 585–588. [CrossRef]
8. Lu, S.; Dai, W.; Tang, Y.; Guo, M. A review of the impact of hydropower reservoirs on global climate change. *Sci. Total Environ.* **2020**, *711*, 134996. [CrossRef]
9. Marques, C.A.F.; Ferreira, J.A.; Rocha, A.; Castanheira, J.M.; Melo-Gonçalves, P.; Vaz, N.; Dias, J.M. Singular spectrum analysis and forecasting of hydrological time series. *Phys. Chem. Earth Parts A/B/C* **2006**, *31*, 1172–1179. [CrossRef]
10. Karran, D.J.; Morin, E.; Adamowski, J. Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes. *J. Hydroinform.* **2014**, *16*, 671–689. [CrossRef]
11. Wang, Z.Y.; Qiu, J.; Li, F.F. Hybrid models combining EMD/EEMD and ARIMA for Long-term streamflow forecasting. *Water* **2018**, *10*, 853. [CrossRef]
12. Zhang, Z.; Zhang, Q.; Singh, V.P. Univariate streamflow forecasting using commonly used data-driven models: Literature review and case study. *Hydrol. Sci. J.* **2018**, *63*, 1091–1111. [CrossRef]
13. Fatichi, S.; Rimkus, S.; Burlando, P.; Bordoy, R. Does internal climate variability overwhelm climate change signals in streamflow? The upper Po and Rhone basin case studies. *Sci. Total Environ.* **2014**, *493*, 1171–1182. [CrossRef] [PubMed]
14. Adombi, A.V.D.P.; Chesnaux, R.; Boucher, M.A. Theory-guided machine learning applied to hydrogeology—State of the art, opportunities and future challenges. *Hydrogeol. J.* **2021**, *29*, 2671–2683. [CrossRef]
15. Najafzadeh, M.; Oliveto, G. Riprap incipient motion for overtopping flows with machine learning models. *J. Hydroinform.* **2020**, *22*, 749–767. [CrossRef]
16. Kisi, O.; Mirboluki, A.; Naganna, S.R.; Malik, A.; Kuriqi, A.; Mehraein, M. Comparative evaluation of deep learning and machine learning in modelling pan evaporation using limited inputs. *Hydrol. Sci. J.* **2022**, *67*, 1–19. [CrossRef]
17. Wang, W.; Van Gelder, P.H.; Vrijling, J.K.; Ma, J. Forecasting daily streamflow using hybrid ANN models. *J. Hydrol.* **2006**, *324*, 383–399. [CrossRef]
18. Wu, C.L.; Chau, K.W. Data-driven models for monthly streamflow time series prediction. *Eng. Appl. Artif. Intell.* **2010**, *23*, 1350–1367. [CrossRef]
19. Freire, P.K.D.M.M.; Santos, C.A.G.; da Silva, G.B.L. Analysis of the use of discrete wavelet transforms coupled with ANN for short-term streamflow forecasting. *Appl. Soft Comput.* **2019**, *80*, 494–505. [CrossRef]
20. Li, S.; Zhang, L.; Du, Y.; Zhuang, Y.; Yan, C. Anthropogenic impacts on streamflow-compensated climate change effect in the Hanjiang River Basin, China. *J. Hydrol. Eng.* **2020**, *25*, 04019058. [CrossRef]
21. Malik, A.; Tikhamarine, Y.; Souag-Gamane, D.; Kisi, O.; Pham, Q.B. Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 1755–1773. [CrossRef]
22. Wang, L.; Li, X.; Ma, C.; Bai, Y. Improving the prediction accuracy of monthly streamflow using a data-driven model based on a double-processing strategy. *J. Hydrol.* **2019**, *573*, 733–745. [CrossRef]
23. Ren, K.; Wang, X.; Shi, X.; Qu, J.; Fang, W. Examination and comparison of binary metaheuristic wrapper-based input variable selection for local and global climate information-driven one-step monthly streamflow forecasting. *J. Hydrol.* **2021**, *597*, 126152. [CrossRef]
24. Zhang, H.; Yang, Q.; Shao, J.; Wang, G. Dynamic streamflow simulation via online gradient-boosted regression tree. *J. Hydrol. Eng.* **2019**, *24*, 04019041. [CrossRef]
25. Rice, J.S.; Emanuel, R.E.; Vose, J.M.; Nelson, S.A. Continental US streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics. *Water Resour. Res.* **2015**, *51*, 6262–6275. [CrossRef]
26. Tyralis, H.; Papacharalampous, G.; Langousis, A. Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms. *Neural Comput. Appl.* **2021**, *33*, 3053–3068. [CrossRef]
27. Ni, L.; Wang, D.; Wu, J.; Wang, Y.; Tao, Y.; Zhang, J.; Liu, J. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *J. Hydrol.* **2020**, *586*, 124901. [CrossRef]
28. Sahour, H.; Gholami, V.; Torkaman, J.; Vazifedan, M.; Saeedi, S. Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. *Environ. Earth Sci.* **2021**, *80*, 1–14. [CrossRef]
29. Zhang, D.; Qian, L.; Mao, B.; Huang, C.; Huang, B.; Si, Y. A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access* **2018**, *6*, 21020–21031. [CrossRef]
30. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363.
31. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the 32nd Conference on Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018.
32. Niu, D.; Diao, L.; Zang, Z.; Che, H.; Zhang, T.; Chen, X. A machine-learning approach combining wavelet packet denoising with Catboost for weather forecasting. *Atmosphere* **2021**, *12*, 1618. [CrossRef]
33. Huang, G.; Wu, L.; Ma, X.; Zhang, W.; Fan, J.; Yu, X.; Zhou, H. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* **2019**, *574*, 1029–1041. [CrossRef]
34. CatBoost. Available online: https://catboost.ai/ (accessed on 30 October 2022).
35. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for big data: An interdisciplinary review. *J. Big Data* **2020**, *7*, 1–45. [CrossRef] [PubMed]
36. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
37. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

38. Brownlee, J. *XGBoost with Python: Gradient Boosted Trees with XGBoost and Scikit-Learn*; Association for Computing Machinery: New York, NY, USA, 2016.
39. Chen, R.C.; Caraka, R.E.; Arnita, N.E.G.; Pomalingo, S.; Rachman, A.; Toharudin, T.; Pardamean, B. An end to end of scalable tree boosting system. *Sylwan* **2020**, *164*, 1–11.
40. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
41. Pavlov, Y.L. *Random Forests*; De Gruyter: Boston, MA, USA, 2000. Available online: https://www.degruyter.com/document/doi/10.1515/9783110941975/html (accessed on 30 October 2022).
42. Louppe, G. Understanding Random Forests: From Theory to Practice. Ph.D Thesis, University of Liège, Liège, Belgium, 2014.
43. Scornet, E. Learning with Random Forests. Ph.D. Thesis, Université Pierre et Marie Curie, Paris, France, 2015.
44. Sensoy, S.; Demircan, M.; Ulupinar, Y.; Balta, Z. Climate of Turkey. Available online: https://mgm.gov.tr/FILES/genel/makale/31_climateofturkey.pdf (accessed on 30 September 2022).
45. Yang, U.; Sheng-tian, Y.; Ming-yong, C.; Qiu-wen, Z.; Guo-tao, D. The Applicability Analysis of TRMM Precipitation Data in the Yarlung Zangbo River Basin. *J. Nat. Resour.* **2013**, *28*, 1414–1425.
46. Santos, C.A.G.; Brasil Neto, R.M.; da Silva, R.M.; Passos, J.S.D.A. Integrated spatiotemporal trends using TRMM 3B42 data for the Upper São Francisco River basin, Brazil. *Environ. Monit. Assess* **2018**, *190*, 175. [CrossRef]
47. Medhioub, E.; Bouaziz, M.; Achour, H.; Bouaziz, S. Monthly assessment of TRMM 3B43 rainfall data with high-density gauge stations over Tunisia. *Arab. J. Geosci.* **2019**, *12*, 15. [CrossRef]
48. Adnan, R.M.; Liang, Z.; Parmar, K.S.; Soni, K.; Kisi, O. Modeling monthly streamflow in mountainous basin by MARS, GMDHNN and DENFIS using hydroclimatic data. *Neural Comput. Appl.* **2021**, *33*, 2853–2871. [CrossRef]
49. Shi, J.; Guo, J.; Zheng, S. Evaluation of hybrid forecasting approaches for wind speed and power generation time series Renewable and Sustainable. *Energy Rev.* **2012**, *16*, 3471–3480.
50. Zhang, D.; Peng, X.; Pan, K.; Liu, Y. A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine. *Energy Convers. Manag.* **2019**, *180*, 338–357. [CrossRef]
51. Kisi, O. River flow forecasting and estimation using different artificial neural network techniques. *Hydrol. Res.* **2008**, *39*, 27–40. [CrossRef]
52. Sanikhani, H.; Kisi, O. River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. *Water Resour Manag.* **2012**, *26*, 1715–1729. [CrossRef]