Article

Assessing the Predictability of an Improved ANFIS Model for Monthly Streamflow Using Lagged Climate Indices as Predictors

Mohammad Ehteram [1], Haitham Abdulmohsin Afan [2,*] , Mojgan Dianatikhah [1], Ali Najah Ahmed [3] , Chow Ming Fai [3], Md Shabbir Hossain [4] , Mohammed Falah Allawi [5] and Ahmed Elshafie [2]

[1] Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan 35131-19111, Iran; mohammdehteram@semnan.ac.ir (M.E.); mojgan_diyanat1370@yahoo.com (M.D.)

[2] Department of Civil Engineering, Faculty of Engineering, University Malaya, Kuala Lumpur 50603, Malaysia; elshafie@um.edu.my

[3] Institute of Energy Infrastructure (IEI), Universiti Tenaga Nasional, 43000 Selangor, Malaysia; Mahfoodh@uniten.edu.my (A.N.A.); Chowmf@uniten.edu.my (C.M.F.)

[4] School of Energy, Geoscience, Infrastructure and Society. Department of Civil Engineering, Heriot-Watt University, Putrajaya 62200, Malaysia; m.hossain@hw.ac.uk or realism007@gmail.com

[5] Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Selangor 43600, Malaysia; mohmmd.falah@gmail.com

* Correspondence: haitham.afan@gmail.com

check for updates

Abstract: The current study investigates the effect of a large climate index, such as NINO3, NINO3.4, NINO4 and PDO, on the monthly stream flow in the Aydoughmoush basin (Iran) based on an improved Adaptive Neuro Fuzzy Inference System (ANFIS) during 1987–2007. The bat algorithm (BA), particle swarm optimization (PSO) and genetic algorithm (GA) were used to obtain the ANFIS parameter for the best ANFIS structure. Principal component analysis (PCA) and Varex rotation were used to decrease the number of effective components needed for the streamflow simulation. The results showed that the large climate index with six-month lag times had the best performance, and three components (PCA1, PCA2 and PCA3) were used to simulate the monthly streamflow. The results indicated that the ANFIS-BA had better results than the ANFIS-PSO and ANFIS-GA, with a root mean square error (*RMSE*) 25% and 30% less than the ANFIS-PSO and ANFIS-GA, respectively. In addition, the linear error in probability space (LEPS) score for the ANFIS-BA, based on the average values for the different months, was less than the ANFIS-PSO and ANFIS-GA. Furthermore, the uncertainty values for the different ANFIS models were used and the results indicated that the monthly simulated streamflow by the ANFIS was computed well at the 95% confidence level. It can be seen that the average streamflow for the summer season is 75 m^3/s, so that the stream flow for summer, based on climate indexes, is more than that in other seasons.

Keywords: ANFIS-BA; ANFIS-PSO; ANFIS-GA; large climate index; ENSO

1. Introduction

The regional understanding of the interrelationship between the catchment attributes and the catchment hydrologic responses could be one of the basic concepts to predict the hydrologic variables for any ungauged catchment [1]. In addition, it should be noticed that the scale of the catchment is one of the factors affecting the catchment's hydrologic response and mainly streamflow. In fact,

streamflow is considered one of the most important hydrological variables that is essential, especially for ungauged catchments. For a catchment which is not gauged for streamflow, implementing a proper interrelationship is considered a real example of the needs for such regionalization methodology and could be very valuable for a few motives [2]. For example, during the design and construction of any hydrological or hydraulic structure, e.g., dams, barrages or bridges, several hydrological parameters might need to be measured before the design or construction. All of these hydrological or hydraulic structures may require a forecasting/prediction/estimation to be carried out for streamflow, as an example of the hydrological variables at ungauged points. If the catchment under consideration is not gauged for streamflow, these estimates must be based on some form of regionalisation, where the catchment is considered to behave similarly to another catchment (or catchments) with similar climatology and landscape attributes [3]. However, this concept does not need to be successful every time, as in a few cases and due to a very sensitive catchment feature, the similarity in the hydrologic response is not similar at all. For most streams, especially those with small ungauged catchments, no record of streamflow is available. In that case, it is possible to make streamflow prediction using the rational method or some modified version of it. However, if chronological (historical) records of streamflow are available, a short-term prediction of the streamflow could be made for a given ungauged point using advanced data-driven models. In this context, there is a need to utilize a new concept of modelling such as machine learning to be able to accurately predict the streamflow [4].

Predications of hydrological variables are considered to be important issues for water managers. The construction of hydraulic structures, or water resource management for a basin, needs accurate prediction of the hydrological variables [5]. Streamflow predication is an important issue for researchers, who attempt to use the best tools, such as different software or hydrological models, to accurately estimate the stream flow over different periods [5,6]. The streamflow predication can be converted to a more complex issue when climate variability has an important effect on the streamflow estimation. Climate variability can convert the streamflow predication into a real challenge [7]. Current oceanic–atmospheric models can account for climate variability over different years. The Pacific Decadal Oscillation (PDO) and the El Nino Southern Oscillation (ENSO) are two important oceanic–atmospheric indices that occur due to sea surface temperature (SST) [8]. SST and sea level pressure (SLP) are two important components of the ENSO event. Anomalous SSTs can be seen in three regions: $NINO_3$ (5° S–5° N, 150°–90° W), NINO3.4 (5° S–5° N, 170°–120° W), NINO4 (5° S–5° N, 150°–160° W) and NINO1 + 2 (0–10° S, 90°–80° W) (Figure 1).
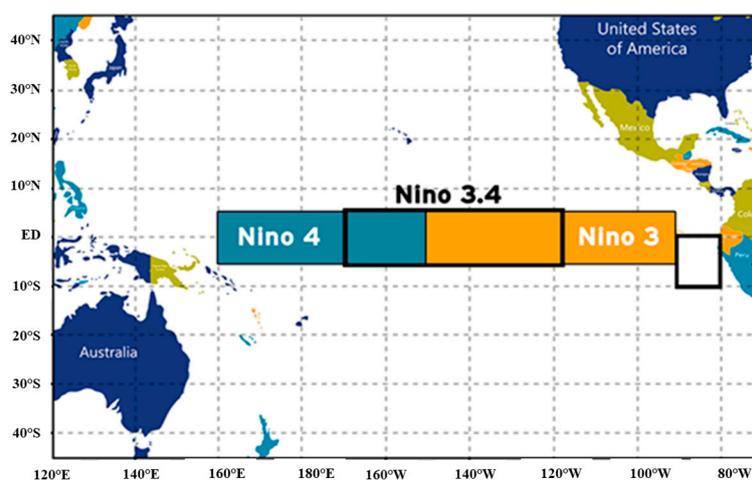

Figure 1. The presentation of ENSO regions.

The ENSO can change the global atmosphere circulation by variation of temperature and precipitation. There is a known interaction between the atmosphere and ocean in the tropical Pacific so that the dry or wet condition and periodic variation between below normal and above normal

SSTs can be seen for this event [9]. When the ENSO occurs, powerful winds are weakened when they are transferred from the eastern to the western side of the Pacific, and the warm equatorial waters are moved to the eastern Pacific and northern South America. However, the ENSO can change the temperature and precipitation and thus, the streamflow volume [10]. The PDO is determined by the SST in the North Pacific Ocean, and the positive or negative phase of the PDO shows a lower or upper SST, respectively, in the central Pacific region [11]. The PDO and ENSO, based on the variation of temperature and precipitation, can change the streamflow, and thus, accurate prediction of the streamflow, based on ENSO and PDO inputs, is considered a complex and important issue [12]. The predication of streamflow, based on the ENSO and PDO models, depends on finding the correlation value between the models and the streamflow and the indexes of ENSO and PDO. Of course, finding the lag times of the used climate indexes with the streamflow for the different years or months is another important issue [5,12]. The predication models need a large amount of data for the streamflow simulation, and thus, artificial intelligence models, such as neural networks or fuzzy models, can be considered as good selections for the streamflow simulation. These models can receive a large amount of data with a high learning ability for a low computational time, so that a high flexibility can be seen between models and the hydrological condition of different basins [13–16]. The present study develops the adaptive neuro-fuzzy inference system (ANFIS) for the predication of streamflow for a case study in Iran, where the ANFIS is improved with the bat algorithm, particle swarm algorithm and genetic algorithm. The hybrid model helps determine the accurate structure of ANFIS models, such as nonlinear and linear parameters in the ANFIS, and then all used models are compared based on the error index. If the unknown values of ANFIS parameters are obtained accurately, the model does not have an accurate predication capability, and thus, the optimal values for the ANFIS parameters can be obtained based on optimization algorithms.

The evaluation of models in different climates is considered to include a comprehensive evaluation of the results. NINO3, NINO3.4, NINO4 and PDO are used as climate indexes in the current paper. In addition, the uncertainty of different used models based on climate indexes is computed for the streamflow predication, whereas previous research simulated the streamflow based on lead times and climate indexes only. Therefore, the current paper includes a comprehensive study of the streamflow predication under the index climate based on artificial intelligence models. Thus, the current study presents a new version of ANFIS that can be used for complex climate events that receive a large amount of data.

2. Background

Kashid et al. [13] used genetic programming for a case study of the predication of streamflow with consideration to the ENSO and the equatorial indicant ocean oscillation (EQUINOO). The results indicated that the weekly streamflow predication based on the ENSO and EQUINOO, with consideration to the current time, had a better performance than the climate indexes with lag times. In their study, the error indexes, such as root mean square error, for the genetic programming based on current time had a lower value than regression models and the used climate index with lag times.

Maity and Kashid [14] simulated weakly streamflow based on the input data of the ENSO, local outgoing radiation (LOR) and previous streamflow information based on genetic programming. The results indicated that the ENSO, LOR and previous streamflow information could improve the accuracy of weak streamflow predication. Different combinations of inputs with different lag times were used for the study.

Classification, regression tree and genetic programming were used for the streamflow predication based on the ENSO, PDO and North Atlantic oscillation (NAO) [17]. The seasonal streamflow was predicated and the results indicated that an accurate predication based on genetic programming for winter and spring could be obtained based on the NAO index with different lag times, and the results of genetic programming had a better correlation with the observed data compared to the other models.

A periodic auto regressive (PAR) model was used for the monthly streamflow predication [18]. The NINO3 index was used for this predication. The cross validation indicated that the PAR, based on the NINO3 data, could enhance the predications with a 3-month lead time. The correlation results showed that the climate indexes were dependent on the monthly streamflow with consideration of lag time.

The Atlantic multidecadal oscillation (AMO), ENSO and PDO were used to simulate the peak season for another study [19]. The least square support vector machine (LSVM) was used for the study and the results indicated that the LSVM could simulate streamflow better than the support vector machine and back propagation neural network, and furthermore, increasing the lead time improved the accuracy of the predications significantly.

The Bayesian neural network (BNN), support vector machine (SVM) and multiple linear regression (MLR) were used to simulate the daily streamflow [20]. The global forecasting system (GFS) outputs plus local observation were used as the best combination of predicators. The results indicated that the BNN based on the NINO3.4 during longer lead times of 5–7 were more accurate than other models based on lower values of the error indexes.

The SVM method was used to simulate spring–summer streamflow for another case study [21]. The North Atlantic Ocean (NAO) was used as the index climate and the results indicated that the best streamflow predications could be obtained with a 6-month, compared to a 3-month or 9-month, lead time.

NAO, SST and ENSO data were used to predicate the annual streamflow for the Colorado River [18]. The results of the SVM were found to be better than the stream predication with a 1-year lead time compared to the back propagation neural network and multiple linear regression. The predications based on the NAO and SST indexes matched the observed data well.

The wavelet–SVM method was used to predict the monthly and daily streamflow based on the ENSO and the Indian Ocean dipole (IOD) [22]. The wavelet–SVM for the monthly (lead times of 1–3 months) and daily (lead times of 1–7 days) streamflow predications had better results than the ANFIS model.

The extreme learning machine (ELM) and artificial neural network (ANN), based on predictors of rainfall, NINO3.SST, NINO4.SST, southern oscillation index (SOI) and IOD, were used to predicate mean streamflow water level [23]. The correlation showed that the best inputs were rainfall, NINO3.SST, NINO4.SST and SOI for the streamflow simulation, and the ELM model was more accurate than the ANN based on lower values for the different indexes.

The SVM and a hydrologic uncertainty processor (HUP) were used for the monthly runoff predication with consideration to the SST index [24]. The HUP could not quantify the simulating reliability but it could generate effective information. The SVM based on a lead time of 1 year could predicate the runoff with high accuracy.

A multiple linear regression was used to predicate long-term streamflow based on the IOD, POD and ENSO indexes in Australia [25]. The correlation coefficient for the streamflow and indexes were used to determine the best input combination and the results indicated that the POD and ENSO based on a one-month lead time could improve the *RMSE* and mean absolute error (MAE) significantly and the predicated streamflow for spring was more accurate.

Another study simulated inflow to three reservoir dams based on the ANFIS-auto regressive exogenous (ARX), ANN-ARX and random forest (RF)-ARX models, and on the NINO1 + 2 and Atlantic meridional mode (AMO) indexes [26]. The results indicated that the indexes based on the ANFIS-ARX model with a 12-month lead time could simulate the streamflow more accurately than the ANN-ARX and RF-ARX models.

However, the artificial intelligence models and regression models based on a large climate index have a wide application in the predication of hydrological variables, such as rainfall, streamflow, drought and ground water level [27–30], and thus, these models based on climate indexes are known to be effective tools for climate studies.

3. Materials and Methods

3.1. ANFIS

Fuzzy logic and ANN combine to form the ANFIS model, which is known as a multi-layer feed forward network. A first order of the Sugeno fuzzy model included the following equations [31]:

$$Rule1 : if(x)isA_1and(y)is(B_1)then(f_1) = p_1x + q_1y + r_1$$
$$Rule2 = f(x)isA_1and(y)is(B_2)then(f_2) = p_2x + q_2y + r_2 \tag{1}$$

where A_1, B_1, B_2 and B_2 are membership functions, x and y are inputs, f_1 and f_2 are output functions and p_1, q_1, r_1, p_2, q_2 and r_2 are linear parameters (Figure 2).
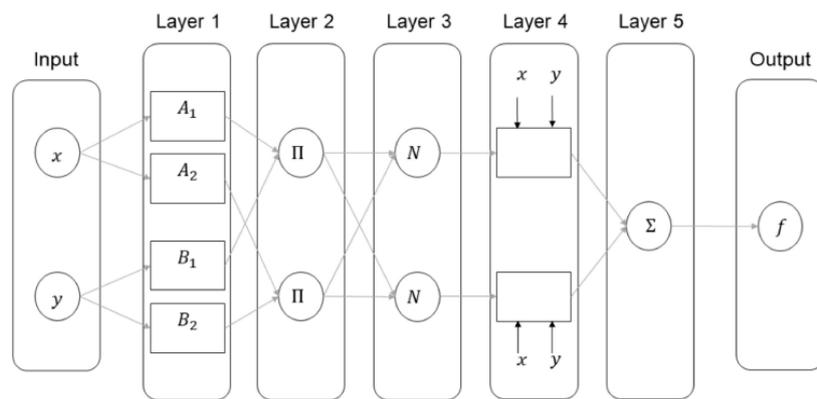

Figure 2. Structure of the ANFIS model.

Layer 1: the fuzzy membership function (MF) generates the fuzzy membership grads for the nodes. The MF converts each point in the input space to a membership with a value between 0 and 1. The outputs of the fuzzification layers are shown based on the following equations:

$$Q_{1,i} = \mu_{Ai}(x), i = 1,2 \tag{2}$$

$$Q_{1,i} = \mu_{B_{i-2}}(y), i = 3,4 \tag{3}$$

where A_i and B_i are fuzzy sets, $\mu_{Ai}(x)$ and $\mu_{B_{i-2}}(y)$ are the degrees of MF, and x and y are the inputs for node i.

Layer 2: the firing strength is shown by the outputs and they are generated by the incoming signals:

$$Q_{2i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(x), i = 1,2 \tag{4}$$

Layer 3: the normalized firing strength is shown by the following equation:

$$Q_{3i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, i = 1,2 \tag{5}$$

Layer 4: the contribution of the ith rule is computed based on the consequent parameters (p_i, q_i, r_i):

$$Q_{4,i} = \overline{w}f_i = \overline{w}_i(p_ix + q_iy + r_i) \tag{6}$$

Layer 5: finally, the summation of rules for each signal node is computed to obtain the outputs:

$$Q_{5,i} = \sum_i \overline{w}f_i = \frac{\sum_i w_if_i}{\sum_i w_i} \tag{7}$$

The best selections for the shape factions in the previous study were the normalized Gaussian and bell-shaped MF. The Gaussian MF for this study was selected because it is smooth and non-zero for all points [31]:

$$\mu_{A_i}(x) = \left(-\frac{(x - c_i)^2}{2\sigma_i 2}\right) \tag{8}$$

where c_i and σ_i are the parameters for the membership function.

The current study attempts to improve the ANFIS structure based on accurate determination of the optimal values of the linear and membership function parameters. Thus, an initial guess of these parameters was inserted into the optimization algorithms and the guesses are known as decision variables. In fact, they were considered as an initial population for the algorithms, and then an objective function, such as *RMSE* [11], was defined for the optimization algorithm. Therefore, the optimization process used for each algorithm could give the optimal values of the parameters for the ANFIS structure.

3.2. Bat Algorithm (BA)

The Bat algorithm is known to be an effective tool for optimization problems. Previous research has shown that it is highly capable of dealing with different issues such as water resource management, energy generation and nonlinear mathematical functions [32,33]. The bats can differentiate the obstacles from food based on sound. In fact, they generate a loud sound and then receive sound echoes at a specific frequency. Three main assumptions can be made for the BA [33]:

(1)   All bats use echolocation to identify the food location.
(2)   The bats fly at the random velocity (v_l) at the location y_l with the frequency $f_{\min}$ and the wavelength λ_l. The loudness parameter for the bats is given by A_0.
(3)   The volume can vary from A_0 to $A_{\min}$.

The velocity and position for each bat is computed based on following equations:

$$f_l = f_{\min} + (f_{\max} - f_{\min}) \times \beta \tag{9}$$

$$v_l(t) = [y_l(t - 1) - Y_*] \times f_L(t) \tag{10}$$

$$y_l(t) = y_l(t - 1) + v_l(t) \tag{11}$$

where $y_l(t - 1)$ is the position at time $(t - 1)$, β is a random value, $f_{\max}$ is the maximum frequency, $f_{\min}$ is the minimum frequency, f_l is the frequency at each iteration, Y_* is the best location for the bats, $y_l(t)$ is the position at time t and $v_l(t)$ is velocity at time t.

A random walk is used as a local search algorithm for the BA:

$$y_t = y(t - 1) + \varepsilon A(t) \tag{12}$$

where ε is a random value between -1 and 1, and $A(t)$ is the volume of the sound.

The volume and pulsation rate (r_l) are updated for each level. The pulsation rate increases and volume decrease when the bats find the food. The volume and pulsation rate are updated based on following equation:

$$r_l^{t+1} = r_l^0[1 - \exp(-\gamma t)]A_l^{t+1} = \alpha A_l^t \tag{13}$$

where γ and α are the constant values (Figure 3).
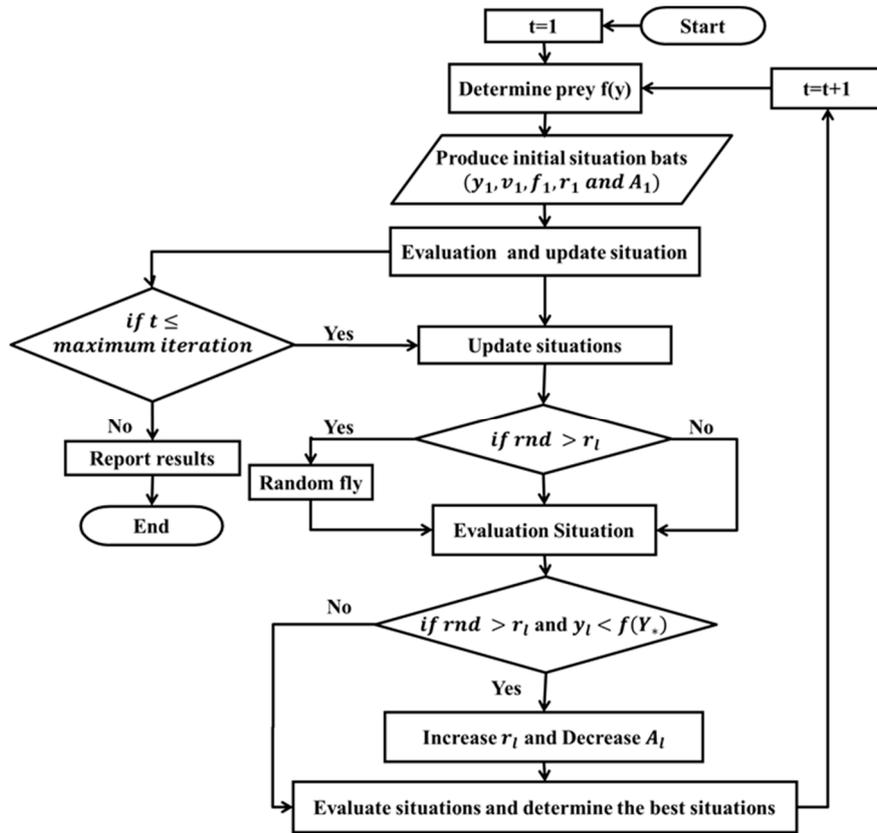
Figure 3. The structure of BA (rnd: random number).

3.3. Particle Swarm Optimization (PSO)

The PSO algorithm is known to have instant convergence, a simple structure and high flexibility for solving complex nonlinear problems [34]. The particle positions are considered as decision variables and the particles attempt to find the best position. First, the algorithm considers the initial positions, $(P(k))$, and in this way, the particle's $x_{is}(k)$ position $(P_i \in P_k)$ equals $(k = 0, k$: number of levels), which is known as the first step. Each particle's F function is computed based on the following equation:

$$if(F(x_i(k))) < pbest_i \rightarrow then \begin{bmatrix} p_{best_i} = F\left(\left(x_i^k\right)\right) \\ x_{pbesti} = x_i(k) \end{bmatrix} \tag{14}$$

where p_{best_i} is the best position of the ith particle.

Equation (14) is used to examine the optimal efficiency of individual particles:

$$if(F(x_i(k))) < gbest_i \rightarrow then \begin{bmatrix} g_{best_i} = F\left(\left(x_i^k\right)\right) \\ x_{gbesti} = x_i(k) \end{bmatrix} \tag{15}$$

where g_{best_i} is the best global position obtained by the different particle swarms.

Then, the velocity for each particle is computed based on the following equation:

$$v_i^k = w v_i^{k-1} + r_1 c_1 \left(x_{pbesti} - x_i^k\right) + r_2 c_2 \left(x_{gbesti} - x_i^k\right) \tag{16}$$

where r_1 and r_2 are random parameters, w is the inertia weight and c_1 and c_2 are the acceleration coefficients.

3.4. Genetic Algorithm (GA)

The GA is known to be a useful algorithm for nonlinear, stochastic and complex problems [35]. First, an initial population is considered for the GA, and the chromosomes are considered as the initial population. Then, an objective function value for each member should be computed to generate the next generation. The next generation is produced based on a selection operator in the reproduction process so that the best chromosomes with the best objective function from current generation values are selected to produce the next generation. The chromosomes with the best objective function value have the highest chance for selection and production of the next generation. Finally, the crossover operator is used to produce the child chromosomes from two different parent chromosomes.

The linear and nonlinear ANFIS parameters were inserted as the initial population into the different algorithms, and then the train level for the ANFIS was simulated and an objective function, such as an error index (*RMSE*), was considered to evaluate the parameter values. Then, the top criteria were checked and if not satisfactory, the algorithm was inserted into an optimization level based on the optimization process in the evolutionary algorithms. Finally, the test level was considered for the data. More detail is provided in Figure 4.
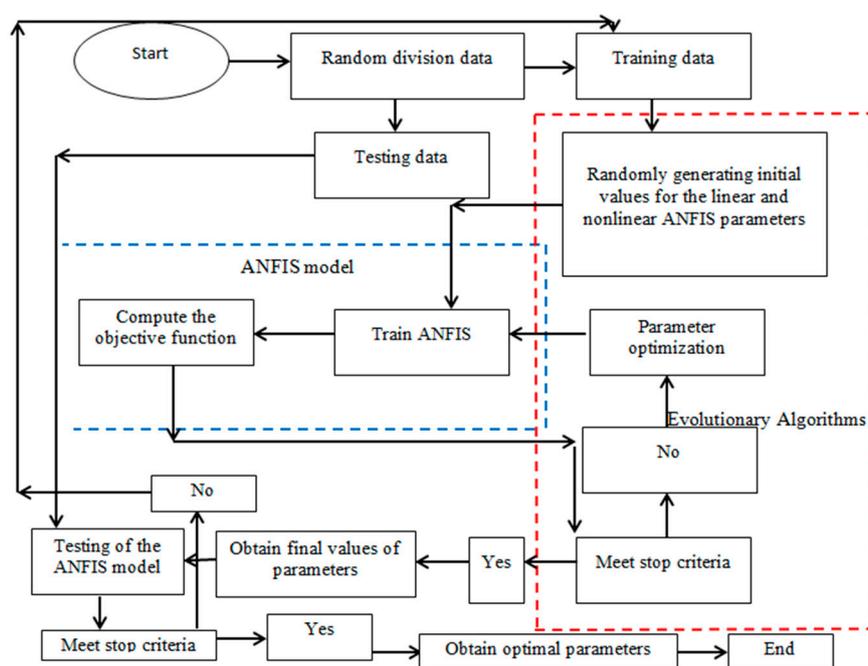

Figure 4. ANFIS and Evolutionary Algorithm [26].

3.5. Principal Component Analysis (PCA)

The effective inputs for the streamflow prediction, based on the climate index, were identified based on principal component analysis (PCA). PCA is an effective method when some input variables should be used for the prediction of hydrological variables, such as streamflow. PCA converts initial variables to new components and thus, these components can be used instead of the initial variables [34]. Furthermore, when all the variables are used to generate the new components, the low-level information may be lost based on this conversion. PCA is based on the following levels:

(1) PCA is considered to be a statistical nonparametric method and thus, it is necessary to evaluate the Kaiser–Meyer–Olkin (KMO) test. This index is computed based on simple and partial correlation coefficients. If the value of the KMO coefficient is more than 0.5, the PCA method can be applied to the data [36–38].

$$KMO = \frac{\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{p} r_{ij}^2}{\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{p} r_{ij}^2 + \sum\limits_{i=1}^{p}\sum\limits_{j=1}^{p} r_{ij}^2} i \neq j \tag{17}$$

where r_{ij}^2 and a_{ij} are the simple correlation coefficient and partial correlation coefficient, respectively, between variables *i,j*.

(2)　The second level is used for the conversion of data to the standard format:

$$Z = \frac{X - \mu}{\sigma} \tag{18}$$

where *Z* is the standard value for the data, μ is the average of each variable and σ is the standard deviation for each variable.

(3)　The correlation matrix is computed to show the variations in the samples and the correlations of different variables with each other. The members of the main diagonal of the matrix are considered as variance of the input variables, and other arrays are considered as covariances of the input variables [34].

(4)　The Eigen vectors and Eigen values are computed based on the following equation:

$$\left| R - \lambda I_p \right| = 0 \tag{19}$$

where *R* is the correlation matrix, λ is the Eigen value, and *I* is the unit matrix. It should be noted that Eigen vectors describe the component characteristics and each component includes a percentage of initial information. *A* higher Eigen value shows that the generated component of the Eigen value includes a higher percentage of initial data. The selection of some initial components based on the highest value of their variance is considered to be important for the PCA.

All initial variables are used to generate the components and thus, the interpretation of the components is difficult for the decision maker. Varimax rotation is considered for the rotation of components and the application of this method means that the number of effective parameters must decrease in order to improve the analysis.

3.6. Data Splitting

One of the first decisions to make when starting a modeling project is how to utilize the existing data. One common technique is to split the data into two groups typically referred to as the training and testing sets and spliting the data is usually for cross-validatory purposes. One portion of the data is used to develop a predictive model in two stages (training and validation) and the other to evaluate the model's performance, which is the testing partition. The training set is used to develop models and feature sets; they are the substrate for estimating parameters, comparing models, and all of the other activities required to reach a final model. The performance of the proposed model is calibrated and validated using the first part of the data before switching to the testing using the test part of the data. The test set is used only at the conclusion of these activities for estimating a final, unbiased assessment of the model's performance

4. Case Study

The considered case study is known as Aidoughmoush and it is located in the northwest of Iran. The Aydoughmoush River is the largest river in this basin. This study considers the streamflow simulation under large-scale climate indexes for the hydrometer station Motorkhaneh in Figure 5. The respective mean yearly precipitation and runoff for this station are 184 mm and 190×10^6 m³ per year. The data from 1987–2007 were available. The river water regulation and the irrigation demand supply for the Aidoughmoush basin were considered and thus, the construction of the Aydoughmoush

dam was also considered. The network area for this basin is 13,500 ha, with 1341.5 ha at a water level above sea level for this dam. Table 1 shows the predicator as input variables and the source of the data collection. There are seven stations for the precipitation measurement, and the Koppen index was used to classify the region's climate [39]. The Koppen classification includes five main climate groups, with the groups classified based on precipitation and temperature (Appendix A). The central part of the basin is shown by the symbol Bwk, which refers to the cold desert climate based on the Koppen classification. The upstream part of the basin is shown by the Cwb symbol, where there is a dry winter and a warm summer, and the downstream part of the basinis shown by Bsk, referring to a cold semi-arid climate. There are some stations that measure precipitation, including Maktu, Ghezel gheye, Tlkhab, Kangavr, Tunnel 7, Poldokhtar and Tazekand. Furthermore, the effect of some stations that are out of the basin were considered because they are located close to the edges and may affect the basin. The inverse distance weighting method was used to obtain the precipitation values for the different points in the maps. The power parameter for this method was obtained based on the optimization algorithm. In fact, the *RMSE* between the observed and simulated precipitation was used as an objective function and then, the power parameter, as an initial population, was inserted into the BA so that the optimization algorithm gave the optimal value for the power parameter. This is because minimizing the *RMSE* is suitable for the decision maker [40].

$$z^* = \frac{\sum\limits_{i=1}^{n} \frac{1}{D_i^q} z_i}{\sum\limits_{i=1}^{n} \frac{1}{D_i^q}} \tag{20}$$

where z^* is the estimated precipitation for each point, D_i is the measured distance between the prediction and observation point, and q is the power parameter. The spatial correlation based on the IDW was 0.94 and Figure 5 shows the spatial distribution of precipitation.
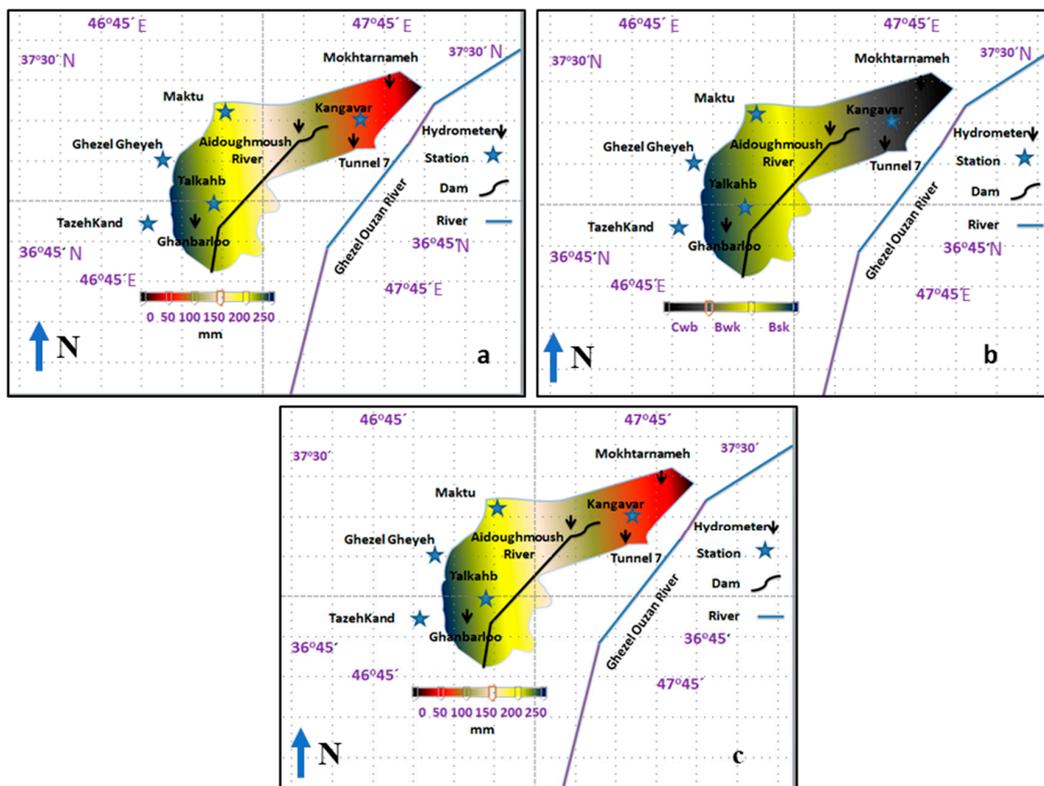

Figure 5. (**a**) The location of the case study, (**b**) The spatial of climate change and (**c**) the spatial distribution based on rainfall (mm).

Table 1. Detail of used climate indexes for the basin.

Predicators	Predicator Definition	Origin	Data Period	Data Source
NINO4	Average SST anomaly over centre Pacific Ocean	Pacific Ocean	1987–2007	https://library.noaa.gov http://sdwebx.worldbank.org/climateporta
NINO3	Average SST anomaly over centre Pacific Ocean	Pacific Ocean	1987–2007	https://library.noaa.gov http://sdwebx.worldbank.org/climateporta
NINO3.4	Average SST anomaly over centre Pacific Ocean	Pacific Ocean	1987–2007	https://library.noaa.gov http://sdwebx.worldbank.org/climateporta
PDO	Average SST anomaly over centre Pacific Ocean	Pacific Ocean	1987–2007	http://research.jisao.washington.edu/pdo/ PDO.latest.txt

In addition, the spatial distribution of precipitation for the total period of 1987–2007 was shown and it is clear that there was more precipitation in the downstream part of the basin or the Bsk climate, and lower precipitation for the Cwb climate in the upstream part of the basin. The monthly streamflow can be predicted based on the large climate index if the effective predictors are identified well.

5. Discussion and Results

5.1. Results of PCA

The current study considers the NINO4, NINO3, NINO3.4 and PDO for a monthly streamflow simulation for the period of 1987–2007. The origin of the initial events for these indexes was significantly distant to the current case study and thus, the effect of these indexes on the streamflow must be considered based on lag times. Thus, the effect of the mentioned indices includes lag times of t, $(t-3)$, $(t-6)$ and $(t-9)$. $t-3$ means that the streamflow at the current time (t) is dependent on the value of the indexes 3, 6 and 9 months prior. Thus, twelve input variables were considered for the current study. Twelve components were considered because there are twelve input variables. The correlation coefficient was based on the 12th order matrix. The KMO for the PCA method is 0.89, which is a good value for the application of the PCA method. Table 2 shows the variation of cumulative covariance and covariance based on percentages for the different components of the PCA. In addition, the value of each variable was computed and it is clear that the first three components had a higher value than the other components. The variance value shows that the first three components had more effect on the streamflow simulation and the first three models included a larger part of information from the initial variables. Furthermore, the Eigen vector value of variable coefficients and the first five components and their coefficients are shown in Table 3. Table 2 shows that the effect of other components was very low based on variance values. In fact, the cumulative variance for the first five components included 90% of the data. For example, the PCA1 can be shown by this equation. Table 2 shows the correlation of different PCA components with the monthly streamflow. The average correlation for the PCA1 was 0.58, while the average value of the correlation coefficient for the other PCA components was less than the average value of PCA1.

$$
\begin{aligned}
PC1 = {}& 0.12NINO3(t) + 0.67NINO3(t-3) + 0.94NINO3(t-6) + 0.61NINO3(t-9) + 0.11NINO4(t) \\
& +0.62NINO4(t-3) + 0.91NINO4(t-6) + 0.60NINO4(t-9) + 0.10NINO3.4(t) + 0.61NINO3.4(t-3) + \\
& 0.89NINO3.4(t-6) + 0.60NINO3.4(t-9) + 0.11PDO(t) + 0.59PDO(t-3) + 0.90PDO(t-6) + 0.62PDO(t-9)
\end{aligned}
\tag{21}
$$

The other components can be shown by the same equation as Equation (20).

**Table 2.** Computed value for each component, the computed comulative varince for each component.

Components	Value of Each Component from 16	Varince Prenatage of Data	Comulative Variantagece Prenatage
PCA1	6.72	42.000	42.000
PCA2	3.68	23.000	65.000
PCA3	2.08	13.000	78.000
PCA4	1.12	7.000	85.000
PCA5	0.88	5.500	90.500
PCA6	0.80	5.000	95.500
PCA7	0.496	3.100	98.600
PCA8	0.179	1.12	99.720
PCA9	0.0128	0.08	99.80
PCA10	0.0128	0.08	99.88
PCA11	0.0064	0.04	99.92
PCA12	0.0064	0.04	99.96
PCA13	0.0016	0.01	99.97
PCA14	0.0016	0.01	99.98
PCA15	0.0016	0.01	99.99
PCA16	0.0016	0.01	100

Table 3. Computed of coefficient for determination of components.

Components	PCA1	PCA2	PCA3	PCA4	PCA5
*NINO*3 (t)	0.12	0.11	0.09	0.07	0.06
*NINO*3 ($t-3$)	0.67	0.64	0.61	0.55	0.53
*NINO*3 ($t-6$)	0.94	0.92	0.91	0.90	0.87
*NINO*3 ($t-9$)	0.61	0.60	0.59	0.52	0.50
*NINO*4 (t)	0.11	0.10	0.08	0.05	0.03
*NINO*4 ($t-3$)	0.62	0.60	0.57	0.55	0.51
*NINO*4 ($t-6$)	0.91	0.90	0.88	0.86	0.85
*NINO*4 ($t-9$)	0.60	0.55	0.52	0.50	0.49
*NINO*3.4 (t)	0.10	0.09	0.08	0.07	0.05
*NINO*3.4 ($t-3$)	0.61	0.56	0.51	0.49	0.42
*NINO*3.4 ($t-6$)	0.89	0.82	0.80	0.79	0.77
*NINO*3.4 ($t-9$)	0.60	0.52	0.49	0.45	0.40
PDO (t)	0.11	0.10	0.09	0.08	0.07
PDO ($t-3$)	0.59	0.57	0.55	0.51	0.45
PDO ($t-6$)	0.90	0.89	0.82	0.80	0.79
PDO ($t-9$)	0.62	0.60	0.57	0.44	0.42

Figure 6 shows the coefficient values for PCA1, PCA2 and PCA3 as the best PCA components. The coefficient values were uploaded based on Table 4 and it is clear that different climate indexes with the lag time ($t-6$) had the most effect compared to other lag times in their group for PCA components. The greatest effects were related to *NINO*3($t-6$), *NINO*4($t-6$), *NINO* 3.4 ($t-6$) and PDO ($t-6$).

However, PCA1–3 as the best components were inserted into the ANFIS models. It should be noticed that 4 components can be enough but 5 components are considered in this study in order to substantiate the reliability of the results and achieve better accuracy.
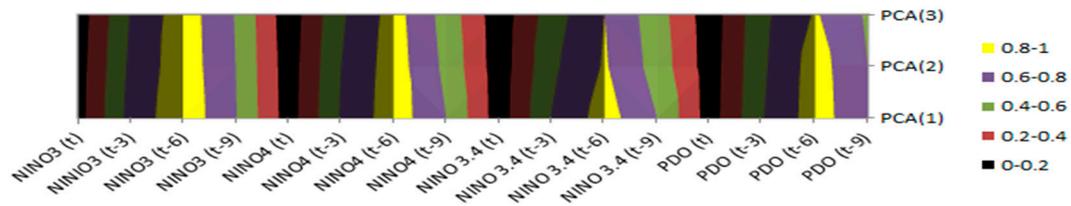
Figure 6. The upload of coefficient values for the different climate indexes and different PCA components.

Table 4. Characteristics of main components based on Varex rotation.

Components	PCA1	PCA2	PCA3	PCA4	PCA5
*NINO*3 (t)	0.12	0.10	0.08	0.06	0.05
*NINO*3 ($t - 3$)	0.56	0.55	0.52	0.49	0.48
*NINO*3 ($t - 6$)	0.91	0.89	0.87	0.86	0.83
*NINO*3 ($t - 9$)	0.55	0.53	0.50	0.49	0.48
*NINO*4 (*t*)	0.11	0.12	0.10	0.09	0.08
*NINO*4 ($t - 3$)	0.52	0.49	0.47	0.45	0.44
*NINO*4 ($t - 6$)	0.90	0.88	0.85	0.83	0.82
*NINO*4 ($t - 9$)	0.50	0.45	0.45	0.42	0.41
*NINO*3.4 (t)	0.09	0.08	0.06	0.05	0.05
*NINO*3.4 ($t - 3$)	0.51	0.50	0.49	0.47	0.45
*NINO*3.4 ($t - 6$)	0.30	0.29	0.27	0.26	0.24
*NINO*3.4 ($t - 9$)	0.50	0.44	0.47	0.45	0.43
PDO (t)	0.08	0.07	0.06	0.06	0.05
PDO ($t - 3$)	0.50	0.47	0.46	0.44	0.42
PDO ($t - 6$)	0.29	0.27	0.25	0.22	0.21
PDO ($t - 9$)	0.49	0.45	0.44	0.40	0.38

5.2. Study of Sensitivity Analysis by Vayring Parameter Values

The optimization algorithms have random parameters that need accurate values obtained for them based on sensitivity analysis. Sensitivity analysis in the current study showed the variation values of parameters versus the objective function values for different population sizes. The current study considered the *RMSE* and the minimization of it as an objective function. For example, the maximum frequency (max*f*) of the BA based on Table 5 varied from 3 to 9 Hz and the best value for this parameter was 7 Hz with consideration given to the population size of 60 because the objective function for the max*f* 7 Hz is 2.20, which is less than other objective function values for the other values of max*f* and population size. The maximum loudness (*A*) varied from 0.30 to 0.9 for the population size of 20 to 80 and the best value for this parameter was 0.7 dB because the lowest value for the objective function occurs when the population size and A are 60 and 0.7 dB, respectively. Other parameters, such as minimum frequency (min*f*), maximum loudness (*A*), mutation probability, crossover probability, acceleration coefficient (c_1 and c_2) and inertia weight (*w*), were calculated as shown in Table 5.

Table 5. Sensitivity analysis for BA, PSO and GA.

BA							
Objective Function	Maximum Load Ness	Objective Function	Minimum Frequency	Objective Function	Maximum Frequency	Objective Function	Population Size
2.6	0.3	2.9	1	3.1	3	2.7	20
2.5	0.5	2.7	2	2.9	5	2.3	40
2.2	0.7	2.2	3	2.2	7	2.2	60
2.7	0.90	2.8	4	3.2	9	2.4	80
PSO							
Objective Function	w	Objective Function	C_2	Objective Function	C_1	Objective Function	Population Size
3.5	0.3	4.4	1.6	4.1	1.6	4.12	20
2.89	0.5	3.1	1.8	3.90	1.8	3.89	40
2.93	0.7	2.2	2.0	3.82	2.0	3.82	60
3.23	0.90	2.8	2.2	3.89	2.2	3.94	80
GA							
	Objective Function	Crossover Rate	Objective Function	Mutation Probability	Objective Function	Population Size	
	7.01	0.30	7.12	0.20	7.25	20	
	6.14	0.50	6.91	0.40	6.92	40	
	6.34	0.70	6.14	0.60	6.12	60	
	6.52	0.90	6.45	0.80	6.25	80	

5.3. Results for Comparison of ANFIS-BA, ANFIS, PSO and ANFIS GA

Table 6 shows the comparison between different results that have been achieved from different ANFIS models. Four evaluation metrics have been used in order to examine the performance for each ANFIS model and in order to select the one that could achieve the highest accuracy and could provide a more consistent accuracy pattern. These evaluation metrics are Root Mean Square Error (*RMSE*), Mean Absolute error (*MAE*), Weightage Index (*WI*), and Nash–Sutcliffe efficiency (*NSE*).

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |X_{obs} - X_{st}| \tag{22}$$

$$RMSE = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^{T} ((X_{obs}) - (X_{st}))^2}}{T} \tag{23}$$

$$NSE = 1 - \frac{\sum_{t=1}^{T} (X_{obs} - X_{st})^2}{\sum_{t=1}^{T} \left(X_{st} - \overline{X_{st}}\right)^2} \quad -\infty \le NSE \le 1 \tag{24}$$

$$WI = 1 - \left[\frac{\sum_{t=1}^{T} (X_{obs} - X_{st})^2}{\sum_{t=1}^{T} \left(\left|X_{st} - \overline{X_{abt}}\right| + \left|X_{st} - \overline{X_{st}}\right|\right)^2} \right] \tag{25}$$

where X_{obs} is the observed data, $\overline{X_{abt}}$ is the average observed data, X_{st} is the simulated data from the model, $\overline{X_{st}}$ is the average simulated data as output from the model, and T is the number of the observed data.

Table 6. Different indexes for evaluation of different ANFIS models (unit for *RMSE* and *MAE*: °C).

Model	Train				Test			
	RMSE	**MAE**	**WI**	**NSE**	**RMSE**	**MAE**	**WI**	**NSE**
ANFIS-GA	3.22	2.89	0.87	0.88	4.25	4.02	0.85	0.84
ANFIS-PSO	3.02	2.85	0.89	0.90	4.01	3.85	0.88	0.86
ANFIS-BA	2.10	1.76	0.95	0.94	2.98	2.78	0.92	0.92

The *RMSE* in the test level for the ANFIS-BA was 25% and 30% less than the ANFIS-PSO and ANFIS-GA, respectively. Such results could be seen for the *RMSE* and train level. The *MAE* error for the ANFIS-BA was 28% and 30% less than the ANFIS-GA and ANFIS-PSO, respectively. The results for the *WI* and *NSE* index show that the ANFIS-BA performed better than the ANFIS-PSO and ANFIS-GA. Although the results indicated that the improved ANFIS achieved the highest accuracy, the error indexes are increased in the testing session. This is due to the fact that the proper sensitivity analysis for optimization algorithms could help improving the results of training and validation sessions and keep iterating until the performance goal is attained, while during testing, the model has to proceed with unseen data and using its structure that has been completed during training and validation session to provide the desired predicted streamflow.
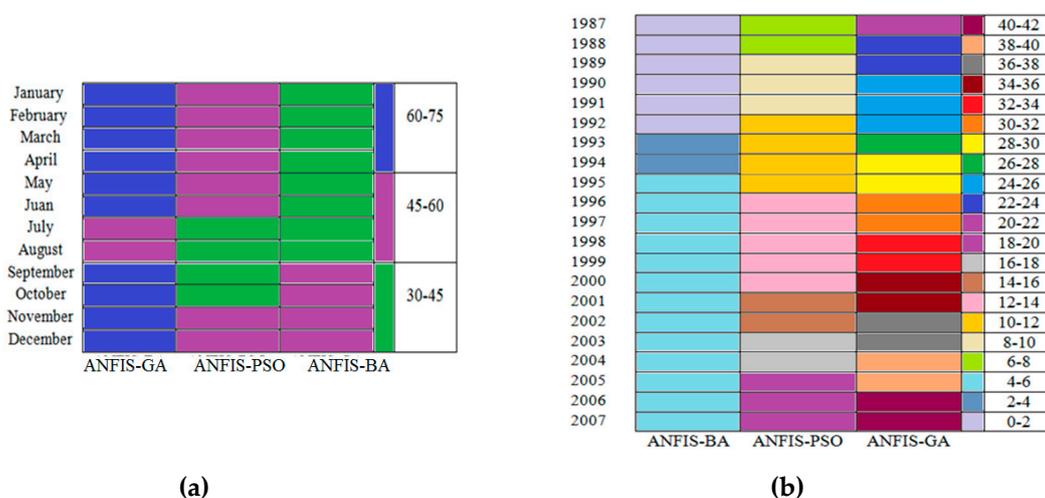
Figure 7 shows the relative error based on percentage for the different years and also the linear error in probability space (LEPS) score based on the average value for the different months during 1987–2007, which was computed to give a better comparison. The difference between observed and simulated values was computed based on the following index [36]:

$$S = 3\left(1 - p_F - p_v + p_f^2 - p_f + p_v^2 - p_v\right) - 1 \tag{26}$$

where p_f is the cumulative probability for the forecasted variable and p_v is the cumulative probability for the observed variable. The probability value for each parameter was computed based on historical probable distribution. The sum of the *S* values was computed based on the following equation:

$$S'' = \sum_{i=1}^{n} S_k \tag{27}$$

where *n* is the number of total prediction and *k* is the number of each prediction level.

(a)

(b)

Figure 7. (**a**) relative error based on percentage and (**b**) LEPS score.

The average *LEPS* index was computed based on the following equation:

$$SK = \frac{\sum\limits_{j=1}^{n} 100 S''_{j}}{\sum\limits_{j=1}^{n} S''_{mj}} \tag{28}$$

where S''_{mj} is computed such as S'' but with consideration of the assumption that the best prediction $(p_f = p_v)$ is computed when S'' has a positive value. If it has a negative value, S''_{mj} is computed based on the assumption of the worst prediction. The value of LEPS is between the worst value (−100) and the best value (100). Figure 7 shows the relative error based on percentage, where the percentage relative error between the predicated and observed streamflow based on the ANFIS-BA varied from 0 to 4, while for the GA, it varied from 20 to 42%, and furthermore, the ANFIS-BA performed better than the ANIFS-PSO based on computed indexes. The average LEPS for different months of the 1987–2007 period is shown in Figure 7, where the average LEPS value for the most months varied from 60 to 75 for the ANFIS-BA, which was more than the ANFIS-PSO and ANFIS-GA. However, the different indexes showed the superiority of the ANFIS-BA compared to the other models.

5.4. Discussion

The spatial distribution for streamflow was based on the previous method in the case study section and the computation of weights was based on the bat algorithm. MotorKhaneh, Aidougmoush dam, Tunnel 7, Ghezel gheye and Ghnabarloo were the hydrometric stations. The classification of the spatial distribution of streamflow requires a statistical process and thus, the Kappa coefficient based on the following equation can show the accuracy value for each map [38]:

$$Kappa = \frac{P_0 - P_C}{1 - P_C} \tag{29}$$

where P_C is the relative observed agreement, and P_c is the hypothetical probability. The probabilities for each observer were computed based on the observed data. *Kappa* equals 1 if the rates have complete agreement. Figure 8 shows that the spatial streamflow for the different methods and the *Kappa* for the ANFIS-BA was 0.91, while it was 0.85 and 0.78 for the ANFIS-BA and ANFIS-GA, respectively. Thus, the ANFIS-BA has the most accurate streamflow for the different parts of the basin.

In fact, there are several time increment scales of concern for streamflow prediction, which mainly depend on the definition and/or the target of the streamflow prediction. Usually, the streamflow prediction time increments could be defined on a daily, weekly, monthly, seasonally or yearly basis. For example, when the purpose of the streamflow prediction is for agricultural processes, where the accuracy and timeliness of the prediction are of essential economic importance to the agricultural process, it is recommended to develop the streamflow prediction model with monthly and weekly time increments. This is due to the fact that before planting, during the growth and at the end of the growing season, the quantity of water plays an essential role for the decision of seed planting and farming decisions. For seasonal or yearly time increment scale streamflow prediction, it is required to study large-scale hydro-climatological circulation patterns. Finally, for most of the reservoir water systems, where the water is stored for future reallocations and redistribution for different water needs, smaller time scale increments for streamflow prediction are needed. Usually, daily and weekly time scale increment streamflow prediction are used for reservoir water systems based on the size and the main purpose of the reservoir. Apparently, a monthly time scale increment for streamflow prediction is the most common time increment required for most of the hydrological studies and purposes. In this context, this study focused on the monthly time increment prediction for streamflow.
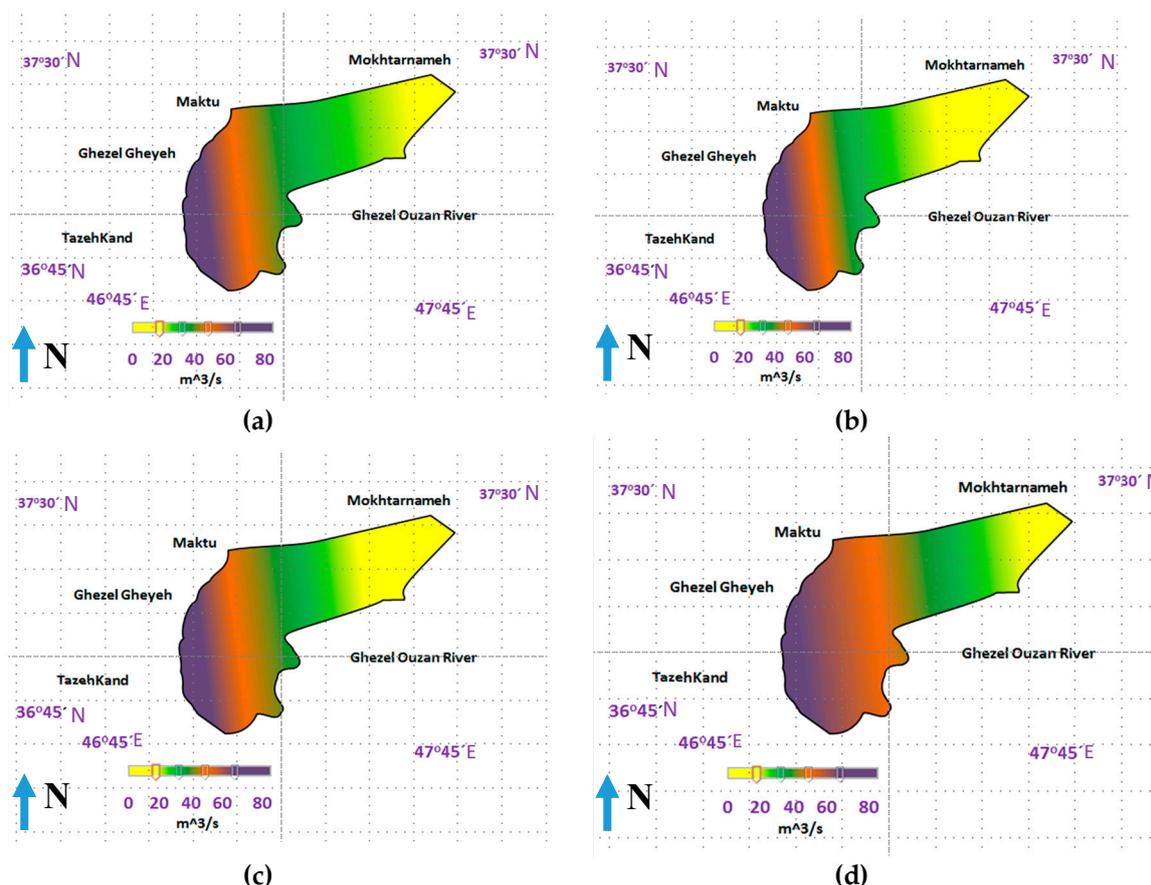
Figure 8. (**a**) spatial streamflow based on ANIFS-BA, (**b**) observed spatial streamflow based on ANFIS-PSO, (**c**) spatial streamflow based on ANFIS-GA, (**d**) spatial streamflow based on observed map.

However, the ANFIS model can be considered as the most appropriate model, but all models could experience uncertainty because all inputs can accord with different levels of uncertainty values. Thus, the uncertainty value for the predicted streamflow in the zones of different maps have been computed. First, 2.5% of the upper and lower domain of simulated streamflow as outlier data was considered and the uncertainty domain was computed based on a 95% confidence level for each predicted point in the different zones of maps. The d factor as an index was used to compare different models based on the division of bound thickness at a 95% confidence level to a standard division of the data. The large value was close to 0, which shows the simulations have a high accuracy. The *p*-value is widely used as a summary statistic of scientific results. The *p*-value is defined as the probability, under the null hypothesis denoting the alternative hypothesis of a population variate, for the variate to be observed as a value equal to or more extreme than the value observed. The *p* factor as the percentage of placement of data at the 95% confidence level was used, with percentages close to 100 indicating better simulations.

Table 7 shows the d factor and p values for the different zones. The p factor for the ANIFS-BA was 90%, while it was 86 and 83% for the ANFIS-PSO and ANIFS-GA, respectively. In addition, the d factor for the ANIFS-BA was 0.52, while it was 073 and 0.75 for the ANIFS-PSO and ANIFS-GA, respectively. However, the general results indicate that the different indexes based on 6-month lag times and PCA were accurate, the ANIFS-BA can simulate the monthly streamflow well, and the decision makers based on PCA1 can understand the effect value of different indexes for the streamflow simulation. Although the uncertainty values for the different models based on the uncertainty value of input data could be effective, the ANIFS-BA simulated the streamflow based on the suitable values for the uncertainty indexes.

Table 7. The computation of uncertainty values for the different models.

Model	*p* Value	*d* Factor
ANFIS-BA	90%	0.52
ANFIS-PSO	86%	0.72
ANFIS-GA	83%	0.75

Finally, the average values for streamflow during different seasons for the 1987–2000 period can be seen in Figure 9. It can be seen that the average streamflow for the summer season is 75 m³/s, which is more than in other seasons. This is related to the snowmelt during summer, which increases the runoff, streamflow and variation of streamflow, and is greater during summer than other seasons. However, the ENSO for the summer can also increase the runoff and streamflow significantly so that the food probability can increase for the summer compared to the other seasons.
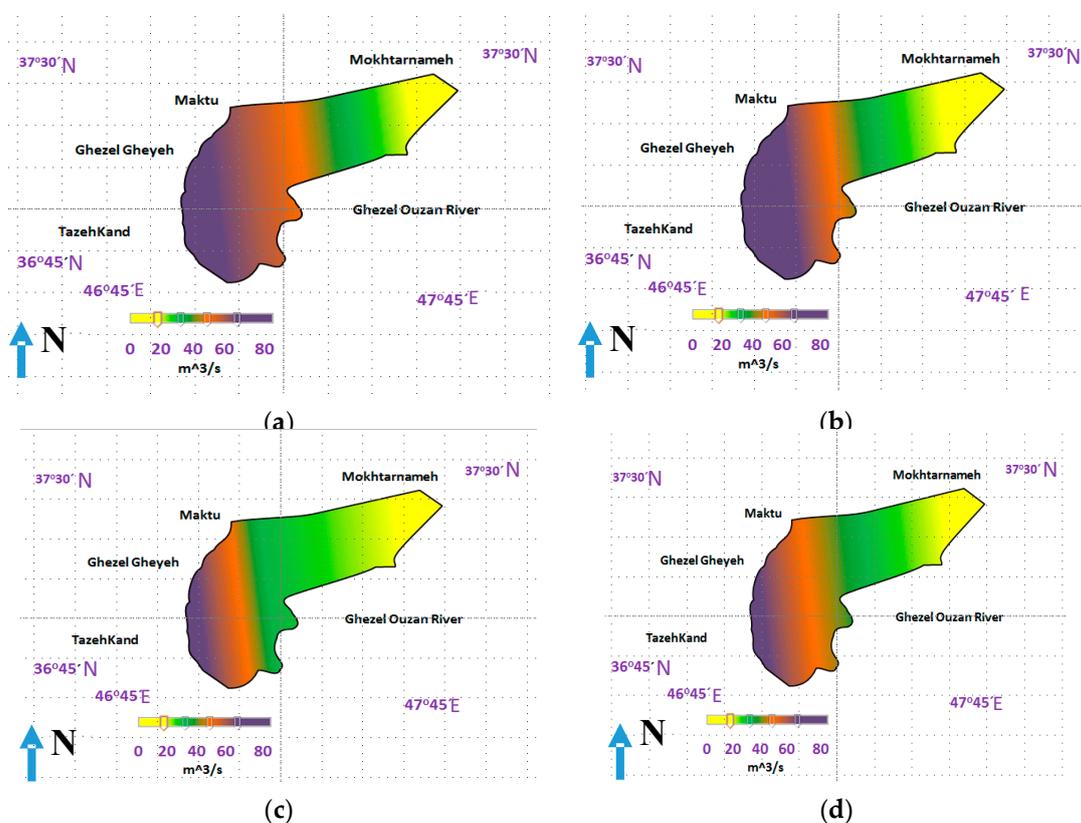
Figure 9. (**a**) spatial streamflow based on summer, (**b**) observed spatial streamflow based on autumn, (**c**) spatial streamflow based on winter, (**d**) spatial streamflow based on spring.

6. Conclusions

The current paper addresses streamflow simulation under large signal climate indexes. The ANFIS method, based on optimization algorithms such as BA, PSO and GA, is designed to obtain the best values for the ANFIS parameter. The Aydoughmoush case study in Iran was used to simulate streamflow with the climate indexes. First, PCA and Varex rotation were used to decrease the component number and identify the components that had an effect on the streamflow. The results indicated that PCA1, PCA2 and PCA3 had the greatest effect and they were therefore inserted into the fuzzy method to simulate the monthly streamflow. The lag time (t-6) also performed well for different indexes such as $NINO_3$, $NINO3.4$, $NINO_4$ and PDO. The results indicated that the ANFIS-BA could decrease the error index more than other methods. For example, the *RMSE* for the ANFIS-BA

was 25 and 30% less than that for the ANFIS-PSO and ANFIS-GA, respectively. In addition, the average LEPS value for the most months varied from 60 to 75 for the ANFIS-BA and was more than the ANFIS-PSO and ANFIS-GA. Also, a weight method was used to obtain the spatial map for the streamflow and *Kappa* coefficient, which had the greatest value for the ANFIS-BA. The results indicated that the ANFIS-BA could have these results because of less uncertainty and the increased summer streamflow due to the snow melt. The current paper showed a large signal climate index could increase or decrease the streamflow and thus, events such as a floods can form through the variations of these indexes. Future studies should consider predicting the streamflow based on satellite images. The results of soft computing methods were compared with such images to determine which tools could produce results with a greater level of agreement compared to the observed data.

Appendix A

First Part	Second Part	Third Part
	W (desert)	-
	S (steppe)	-
B (arid)		h (hot)
		k (cold)
	S (dry summer)	-
	W (dry winter)	-
	F (without dry season)	-
C (temperate)	-	a (hot summer)
	-	b (warm summer)
	-	c (cold summer)

References

1. Goodrich, D.C.; Woolhiser, D.A. Catchment Hydrology. *Rev. Geophys.* **1991**, *29*, 202–209. [CrossRef]
2. Grimaldi, S.; Petroselli, A.; Salvadori, G.; De Michele, C. Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses. *Adv. Water Resour.* **2016**, *90*, 116–133. [CrossRef]
3. Scheel, K.; Morrison, R.R.; Annis, A.; Nardi, F. Understanding the Large-Scale Influence of Levees on Floodplain Connectivity Using a Hydrogeomorphic Approach. *JAWRA J. Am. Water Resour. Assoc.* **2019**, *55*, 413–429. [CrossRef]
4. Dariusz, M.; Andrea, P.; Andrzej, W. Flood frequency analysis by an event-based rainfall-runoff model in selected catchments of southern Poland. *Soil Water Res.* **2018**, *13*, 170–176. [CrossRef]
5. Bhandari, S.; Kalra, A.; Tamaddun, K.; Ahmad, S. Relationship between Ocean-Atmospheric Climate Variables and Regional Streamflow of the Conterminous United States. *Hydrology* **2018**, *5*, 30. [CrossRef]
6. Sulca, J.; Takahashi, K.; Espinoza, J.-C.; Vuille, M.; Lavado-Casimiro, W. Impacts of different ENSO flavors and tropical Pacific convection variability (ITCZ, SPCZ) on austral summer rainfall in South America, with a focus on Peru. *Int. J. Climtol.* **2017**, *38*, 420–435. [CrossRef]
7. Caillouet, L.; Rousseau, A.N.; Savary, S.; Foulon, E. Improving operational ensemble streamflow forecasts by selecting past meteorological scenarios according to climate indices. Proceedings of AGU Fall Meeting, Washington, DC, USA, 10–14 December 2018; Available online: file:///C:/Users/MDPI/Downloads/2018_12_AGU_presentation_Final.pdf (accessed on 15 May 2019).

8. Tamaddun, K.A.; Kalra, A.; Bernardez, M.; Ahmad, S. Effects of ENSO on Temperature, Precipitation, and Potential Evapotranspiration of North India's Monsoon: An Analysis of Trend and Entropy. *Water* **2019**, *11*, 189. [CrossRef]

9. Gomez, F.A.; Lee, S.-K.; Hernandez, F.J.; Chiaverano, L.M.; Muller-Karger, F.E.; Liu, Y.; Lamkin, J.T. ENSO-induced co-variability of Salinity, Plankton Biomass and Coastal Currents in the Northern Gulf of Mexico. *Sci. Rep.* **2019**, *9*, 178. [CrossRef]

10. Tamaddun, K.A.; Kalra, A.; Ahmad, S. Spatiotemporal Variation in the Continental US Streamflow in Association with Large-Scale Climate Signals Across Multiple Spectral Bands. *Water Resour. Manag.* **2019**, *23*, 1947–1968. [CrossRef]

11. Kalra, A.; Sagarika, S.; Ahmad, S. Long-Term Changes in the Continental United States Streamflow and Teleconnections with Oceanic-Atmospheric Indices. *World Environ. Water Resour. Congr.* **2016**, *2016*, 498.

12. Tamaddun, K.A.; Kalra, A.; Ahmad, S. Wavelet analyses of western US streamflow with ENSO and PDO. *J. Water Clim. Chang.* **2016**, *8*, 26–39. [CrossRef]

13. Kashid, S.S.; Ghosh, S.; Maity, R. Streamflow prediction using multi-site rainfall obtained from hydroclimatic teleconnection. *J. Hydrol.* **2010**, *395*, 23–38. [CrossRef]

14. Maity, R.; Kashid, S.S. Short-Term Basin-Scale Streamflow Forecasting Using Large-Scale Coupled Atmospheric–Oceanic Circulation and Local Outgoing Longwave Radiation. *J. Hydrometeorol.* **2010**, *11*, 370–387. [CrossRef]

15. Wei, W.; Watkins, D.W. Data mining methods for hydroclimatic forecasting. *Adv. Water Resour.* **2011**, *34*, 1390–1400. [CrossRef]

16. Thakur, B.; Pathak, P.; Kalra, A.; Ahmad, S. Changing characteristics of streamflow in the Midwest and its relation to oceanic-atmospheric oscillations. Proceedings of AGU Fall Meeting, San Francisco, CA, USA, 12–16 December 2016; Available online: http://adsabs.harvard.edu/abs/2016AGUFM.H33C1549T (accessed on 15 May 2019).

17. Gimenez, J.C.; Lentini, E.J.; Fernández Cirelli, A. Forecasting Streamflows in the San Juan River Basin in Argentina. In *Water and Sustainability in Arid Regions*; Springer: Dordrecht, The Netherlands, 2010; pp. 261–274.

18. Lima, C.H.R.; Lall, U. Climate informed monthly streamflow forecasts for the Brazilian hydropower network using a periodic ridge regression model. *J. Hydrol.* **2010**, *380*, 438–449. [CrossRef]

19. Kalra, A.; Li, L.; Li, X.; Ahmad, S. Improving Streamflow Forecast Lead Time Using Oceanic-Atmospheric Oscillations for Kaidu River Basin, Xinjiang, China. *J. Hydrol. Eng.* **2013**, *18*, 1031–1040. [CrossRef]

20. Rasouli, K.; Hsieh, W.W.; Cannon, A.J. Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J. Hydrol.* **2012**, *414–415*, 284–293. [CrossRef]

21. Kalra, A.; Ahmad, S.; Nayak, A. Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns. *Adv. Water Resour.* **2013**, *53*, 150–162. [CrossRef]

22. Liu, Z.; Zhou, P.; Zhang, Y. A Probabilistic Wavelet–Support Vector Regression Model for Streamflow Forecasting with Rainfall and Climate Information Input*. *J. Hydrometeorol.* **2015**, *16*, 2209–2229. [CrossRef]

23. Deo, R.C.; Şahin, M. Erratum to: An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. *Environ. Monit. Assess.* **2016**, *188*, 90. [CrossRef]

24. Liang, Z.; Li, Y.; Hu, Y.; Li, B.; Wang, J. A data-driven SVR model for long-term runoff prediction and uncertainty analysis based on the Bayesian framework. *Theor. Appl. Climtol.* **2017**, *133*, 137–149. [CrossRef]

25. Esha, R.I.; Imteaz, M.A. Assessing the predictability of MLR models for long-term streamflow using lagged climate indices as predictors: A case study of NSW (Australia). *Hydrol. Res.* **2018**, *50*, 262–281. [CrossRef]

26. Kim, T.; Shin, J.-Y.; Kim, H.; Kim, S.; Heo, J.-H. The Use of Large-Scale Climate Indices in Monthly Reservoir Inflow Forecasting and Its Application on Time Series and Artificial Intelligence Models. *Water* **2019**, *11*, 374. [CrossRef]

27. Zhao, T.; Wang, Q.J.; Schepen, A.; Griffiths, M. Ensemble forecasting of monthly and seasonal reference crop evapotranspiration based on global climate model outputs. *Agric. For. Meteorol.* **2019**, *264*, 114–124. [CrossRef]

28. Reed, E.V.; Cole, J.E.; Lough, J.M.; Thompson, D.; Cantin, N.E. Linking climate variability and growth in coral skeletal records from the Great Barrier Reef. *Coral Reefs* **2018**, *38*, 29–43. [CrossRef]

29. Neves, M.C.; Jerez, S.; Trigo, R.M. The response of piezometric levels in Portugal to NAO, EA, and SCAND climate patterns. *J. Hydrol.* **2019**, *568*, 1105–1117. [CrossRef]

30. Chiri, H.; Abascal, A.J.; Castanedo, S.; Antolínez, J.A.A.; Liu, Y.; Weisberg, R.H.; Medina, R. Statistical simulation of ocean current patterns using autoregressive logistic regression models: A case study in the Gulf of Mexico. *Ocean Model.* **2019**, *136*, 1–12. [CrossRef]

31. Zare, M.; Koch, M. Groundwater level fluctuations simulation and prediction by ANFIS- and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: Application to the Miandarband plain. *J. Hydro-Environ. Res.* **2018**, *18*, 63–76. [CrossRef]

32. Ali, M.; Deo, R.C.; Downs, N.J.; Maraseni, T. Multi-stage hybridized online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm for rainfall forecasting. *Atmos. Res.* **2018**, *213*, 450–464. [CrossRef]

33. Farzin, S.; Singh, V.; Karami, H.; Farahani, N.; Ehteram, M.; Kisi, O.; Allawi, M.; Mohd, N.; El-Shafie, A. Flood Routing in River Reaches Using a Three-Parameter Muskingum Model Coupled with an Improved Bat Algorithm. *Water* **2018**, *10*, 1130. [CrossRef]

34. Chi, S.; Ni, S.; Liu, Z. Back Analysis of the Permeability Coefficient of a High Core Rockfill Dam Based on a RBF Neural Network Optimized Using the PSO Algorithm. *Math. Probl. Eng.* **2015**, *2015*, 1–15. [CrossRef]

35. Montaseri, M.; Hesami Afshar, M.; Bozorg-Haddad, O. Development of Simulation-Optimization Model (MUSIC-GA) for Urban Stormwater Management. *Water Resour. Manag.* **2015**, *29*, 4649–4665. [CrossRef]

36. Solgi, A.; Pourhaghi, A.; Bahmani, R.; Zarei, H. Pre-processing data using wavelet transform and PCA based on support vector regression and gene expression programming for river flow simulation. *J. Earth Syst. Sci.* **2017**, *126*, 65. [CrossRef]

37. Kalra, A.; Ahmad, S. Estimating annual precipitation for the Colorado River Basin using oceanic-atmospheric oscillations. *Water Resour. Res.* **2012**, *48*. [CrossRef]

38. Jiang, Z.; Qi, J.; Su, S.; Zhang, Z.; Wu, J. Water body delineation using index composition and HIS transformation. *Int. J. Remote Sens.* **2011**, *33*, 3402–3421. [CrossRef]

39. Dubreuil, V.; Fante, K.P.; Planchon, O.; Sant'Anna Neto, J.L. Climate change evidence in Brazil from Köppen's climate annual types frequency. *Int. J. Climatol.* **2018**, *39*, 1446–1456. [CrossRef]

40. Tong, W.; Franklin, J.; Zhou, X.; Li, L.; Besenyi, G. Machine Learning on Spark for the Optimal IDW-based Spatiotemporal Interpolation. *Int. Conf. Gisci. Short Pap. Proc.* **2016**. [CrossRef]