

## Article

# A Simple Way to Increase the Prediction Accuracy of Hydrological Processes Using an Artificial Intelligence Model

Ieva Meidute-Kavaliauskiene <sup>1,\*</sup> , Milad Alizadeh Jabehdar <sup>2</sup>, Vida Davidavičienė <sup>1</sup> ,  
Mohammad Ali Ghorbani <sup>2</sup> and Saad Sh. Sammen <sup>3</sup> 

- <sup>1</sup> Department of Business Technologies and Entrepreneurship, Vilnius Gediminas Technical University, 10223 Vilnius, Lithuania; vida.davidaviciene@vilniustech.lt
- <sup>2</sup> Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz 5166616471, Iran; malizadehj97@gmail.com (M.A.J.); ghorbani@tabrizu.ac.ir (M.A.G.)
- <sup>3</sup> Department of Civil Engineering, College of Engineering, Diyala University, Baqubah 32001, Diyala, Iraq; Saad123engineer@yahoo.com
- \* Correspondence: ieva.meidute-kavaliauskiene@vilniustech.lt; Tel.: +370-6-9986847

**Abstract:** Rainfall and evaporation, which are known as two complex and unclear processes in hydrology, are among the key processes in the design and management of water resource projects. The application of artificial intelligence, in comparison with physical and empirical models, can be effective in the face of the complexity of hydrological processes. The present study was prepared with the aim of increasing the accuracy in monthly prediction of rainfall (R) and pan evaporation (EP) by providing a simple solution to determining new inputs for forecasting scenarios. Initially, the prediction of two parameters, R and EP, for the current and one–three lead times, by determining the different input modes, was developed with the SVM model. Then, in order to increase the accuracy of the predictions, the month number ( $\tau$ ) was added to all scenarios in predicting both the R and EP parameters. The results of the intelligent model using several statistical indices (i.e., root mean square error (RMSE), Kling–Gupta (KGE) and correlation coefficient (CC)), with the help of case visual indicators, were compared. The month number ( $\tau$ ) was able to greatly improve the prediction accuracy of both the R and EP parameters under the SVM model and overcome the complexities within these two hydrological processes that the scenarios were not initially able to solve with high accuracy. This is proven in all time steps. According to the RMSE, KGE and CC indices, the highest increase in the forecast accuracy for the upcoming two months of rainfall ( $R_{t+2}$ ) for Ardabil station in scenario 2 (SVM-2) was 19.1, 858 and 125%, and for the current month of pan evaporation ( $EP_t$ ) for Urmia station in scenario 6 (SVM-6), this occurred at the rates of 40.2, 11.1 and 7.6%, respectively. Finally, in order to investigate the characteristic of the month number in the SVM model under special conditions such as considering the highest values of the R and EP time series, it was proved that by using the month number of the SVM model, again, the accuracy could be improved (on average, 17% improvement for rainfall, and 13% for pan evaporation) in almost all time steps. Due to the wide range of effects of the two variables studied in the hydrological discussion, the results of the present study can be useful in agricultural sciences and in water management in general and will help owners.

**Keywords:** rainfall; prediction; pan evaporation; hydrology; artificial intelligence; month number



**Citation:** Meidute-Kavaliauskiene, I.; Jabehdar, M.A.; Davidavičienė, V.; Ghorbani, M.A.; Sammen, S.S. A Simple Way to Increase the Prediction Accuracy of Hydrological Processes Using an Artificial Intelligence Model. *Sustainability* **2021**, *13*, 7752. <https://doi.org/10.3390/su13147752>

Academic Editors: Fadong Li, Gang Chen, Xinlin He and Guang Yang

Received: 10 June 2021  
Accepted: 8 July 2021  
Published: 12 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hydrology is the scientific field that deals with water occurrence, its properties, its distribution and the effect of the atmosphere [1]. It is also related to the design and management of water resource projects. Almost all hydrological activities are unclear because these activities are mainly affected by different interrelated elements [2]. Additionally, hydrological processes are affected by spatial and temporal variability [3]. Therefore, the main challenge that confronts hydrologists is the uncertainty that is considered as the main

feature of hydrological studies. Decision makers and researchers are usually faced with this uncertainty when they try to solve problems related to hydrological events such as floods, evaporation, precipitation, contaminant transport and drought phenomena [2].

There are different approaches that are used in hydrological modeling around the world which include physical models and empirical models. For almost all hydrological studies, physical models were the conventional modeling tools that were used in recent decades [4]. Physically based models are a representation of the ‘physics’ behind a hydrological system and try to solve partial differential equations in order to represent the best understanding of hydrological processes. Physically based models often use two-dimensionally, but sometimes three-dimensionally, distributed data. Due to this data distribution approach, the data required for these models are typically very large. However, the accuracy of these models highly depends on the availability of detailed and accurate data about complex hydrological system properties, which are not usually available due to cost and time limitations, especially in developing countries, resulting in model uncertainties and unsatisfactory performance, which in turn result in insufficient water resource management decisions [5,6]. Empirical models are often used when relations become extremely complicated and difficult to describe. Empirical models are most often utilized in areas with little available information about hydrologic systems. On the downside, empirical models have certain drawbacks concerning their applicability. Based on the above, and due to the aforementioned limitation of using physical and empirical models, it became necessary to find another alternative approach to hydrological modeling [7,8]. Therefore, in recent years, artificial intelligence has attracted considerable attention and has been widely used in modeling different hydrological processes such as rainfall, evaporation, floods and the rainfall–runoff relationship [9–14].

The authors of [15] used an ANFIS (adaptive neuro-fuzzy inference system) in order to predict rainfall in the South Tangerang region, Indonesia. They adopted different input combinations and different membership functions during the training and testing of the model. The authors of [16] presented a comparison study to predict rainfall in Malaysia using different AI techniques that included Bayesian linear regression (BLR), boosted decision tree regression (BDTR), decision forest regression (DFR) and neural network regression (NNR). The authors of [17] adopted an echo state network (ESN), deep echo state network (DeepESN), back-propagation network (BPN) and support vector regression (SVR) to estimate rainfall. The meteorological hourly data from 2002 to 2014 at the Tainan Observatory in southern Taiwan were used to develop the models. The results showed that the correlation coefficient of DeepESN was better than that of ESN, BPN and SVR. On the other hand, the authors of [18] studied the efficiency of some data-driven techniques including support vector regression (SVR) and artificial neural networks (ANN), and combinations of them with wavelet transforms (WSVR and WANN) were investigated for predicting evaporation rates at Tabriz (Iran) and Antalya (Turkey) stations. They used four statistical indices, namely, the root mean square error (RMSE), the mean absolute error (MAE), the correlation coefficient (R) and Nash–Sutcliffe efficiency (NSE), for evaluating the results of modeling. The results indicated that the ANN model performed better than the WANN, SVM and WSVM models for both stations. The authors of [19] built three different AI models based on artificial neural networks (ANNs) (multilayer perceptron (MLP) and radial basis function network (RBFN)) and support vector regression (SVR) in order to estimate the evaporation in Turkey. For the purpose of evaluation, they compared the results of modeling with observed class A pan evaporation data. The outcome showed that the performance of ANN and SVM was similar. The authors of [20] compared radial basis neural network (RBFNN), self-organizing map neural network (SOMNN) and multiple linear regression (MLR) models for prediction of the daily EP in Pantnagar in India. They used the gamma test to choose the input combination. They concluded that the RBFNN model with six input meteorological parameters performed with the highest accuracy compared to the other models. The author of [21] studied the capabilities of a neuro-fuzzy (NF) technique to estimate daily pan evaporation. The results of the NF model were

compared with the results of an ANN model. The results revealed that the proposed NF and ANN models have good abilities to estimate the value of evaporation using different meteorological data. The authors of [22] investigated the accuracy of two heuristic regression approaches, multivariate adaptive regression splines (MARS) and M5 model tree (M5Tree), in estimating pan evaporation using only temperature data as input. The results revealed that the MARS model performed better than the other models.

The overall purpose of the presented study is to increase the accuracy of the developed SVM model for monthly prediction of rainfall (R) and pan evaporation (EP) in different time steps, using a simple and new method, along with input scenarios for the model. In this way, the results of the model in the shadow of this new simple method can be used to predict the monthly rainfall (R) at Ardabil station and pan evaporation (EP) at Urmia station, with high accuracy and for different time steps. The SVM model, unlike other artificial intelligence methods, has a certain mathematical structure, meaning that it does not over-fit during modeling. Due to the lack of long-term data required by SVM, this algorithm contains specific objective and kernel functions and can provide ideal results using optimized mathematical methods [23].

## 2. Methods and Material

### 2.1. Study Region and Datasets

This study was conducted to investigate the capability of the developed support vector machine (SVM) model to predict monthly rainfall (R) at Ardabil station and monthly pan evaporation (EP) at Urmia station, with the effect of the month number ( $\tau$ ) characteristic within the scenarios. For this purpose, these two stations will be introduced first. In order to predict the monthly R and EP, in this study, the monthly rainfall information from Ardabil meteorological station and the EP information from Urmia meteorological station were used. Ardabil and Urmia, as the provincial capitals of Ardabil and West Azerbaijan, are located in Iran, in the mountainous region, and have a semi-arid and cold climate. Additionally, in these areas, rainfall has a disproportionate seasonal distribution. Thus, most of the rainfall, whether in the form of rain or snow, occurs in the autumn–winter and spring seasons. Ardabil meteorological station is located at an altitude of 1335 m above sea level, with a longitude of  $47^{\circ}$  and  $2'$  and latitude of  $38^{\circ}$  and  $13'$ , and Urmia meteorological station has a latitude of  $37^{\circ}$  and  $32'$  and a longitude of  $45^{\circ}$  and  $5'$ , and an elevation of 1316 m. They are both located above sea level. Figure 1 shows a visual view of the geographical location of the meteorological stations under study.

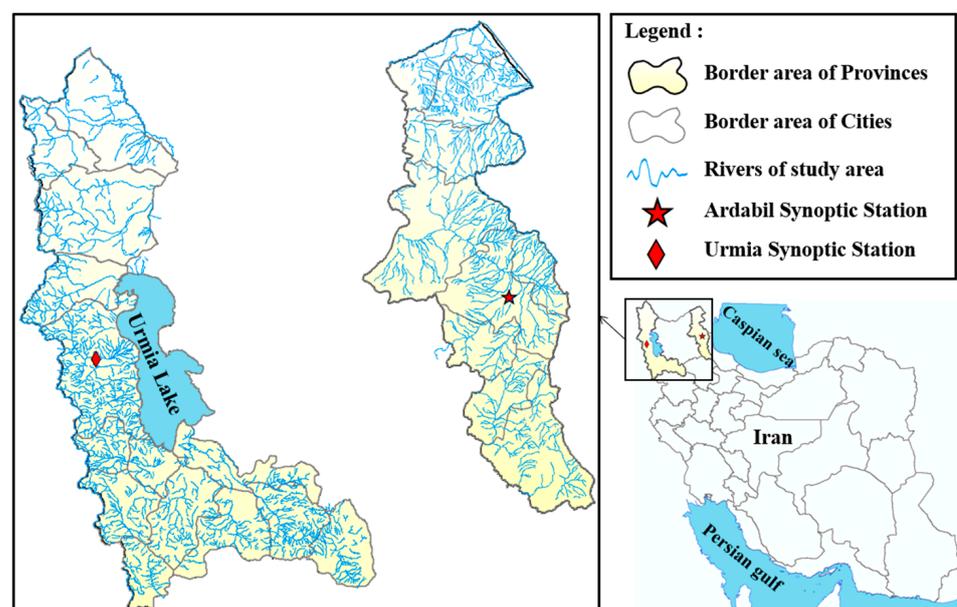
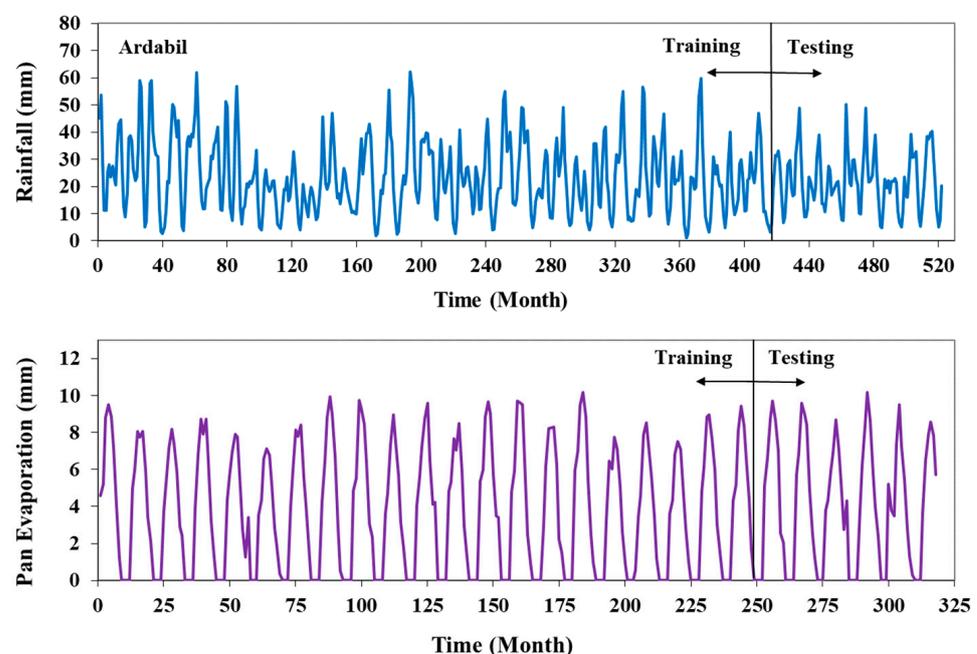


Figure 1. Geographical location of Ardabil and Urmia meteorological stations.

In this study, rainfall (R) data of Ardabil meteorological station, on a monthly basis in the statistical period 1976–2019, and pan evaporation (EP) data of Urmia meteorological station, on a monthly basis in the statistical period 1993–2019, were obtained from the Meteorological Organization. Data were divided so that 80% of the data were isolated for training, and 20% for the testing phase. To be more precise, out of the total 522 months of R data of Ardabil station, 417 months were considered for training (1976–2010), and 105 months for testing (2019–2011). For modeling the EP of Urmia station, out of the total number of 318 months of EP data, 249 months (2013–1993) were considered for training, and 69 months for testing (2019–2014). Figure 2 shows the data segmentation diagram for predicting R and EP at Ardabil and Urmia stations in two sections: training and testing. Table 1 shows the statistical characteristics (minimum, maximum, mean, standard deviation and skewness) at Ardabil and Urmia stations for the training and testing datasets. In order to investigate the different input and output modes of the smart model, cross-correlation analysis was performed between the types of modes. The visual view obtained from the study of the interrelationship of inputs and outputs for the monthly rainfall of Ardabil station and monthly EP of Urmia station is shown in Figure 3.



**Figure 2.** Variation in the monthly rainfall (R) (1976–2019) and pan evaporation (EP) (1993–2019) time series depicting the training and testing datasets at Ardabil and Urmia meteorological stations.

**Table 1.** The monthly statistical parameters of rainfall data at Ardabil station and EP data at Urmia station.

| Station | Datasets | Data No. | * Statistics |       |       |       |          |
|---------|----------|----------|--------------|-------|-------|-------|----------|
|         |          |          | Min          | Max   | Ave   | SD    | Skewness |
| Ardabil | Total    | 519      | 1.02         | 62.25 | 23.19 | 12.92 | 0.63     |
|         | Training | 414      | 1.02         | 62.25 | 23.48 | 13.47 | 0.62     |
|         | Testing  | 105      | 4.85         | 50.07 | 22.05 | 10.47 | 0.44     |
| Urmia   | Total    | 318      | 0            | 10.17 | 3.91  | 3.36  | 0.17     |
|         | Training | 249      | 0            | 10.16 | 3.86  | 3.36  | 0.19     |
|         | Testing  | 69       | 0            | 10.17 | 4.11  | 3.41  | 0.11     |

\* Min, minimum; Max, maximum; Ave, average; SD, standard deviation.

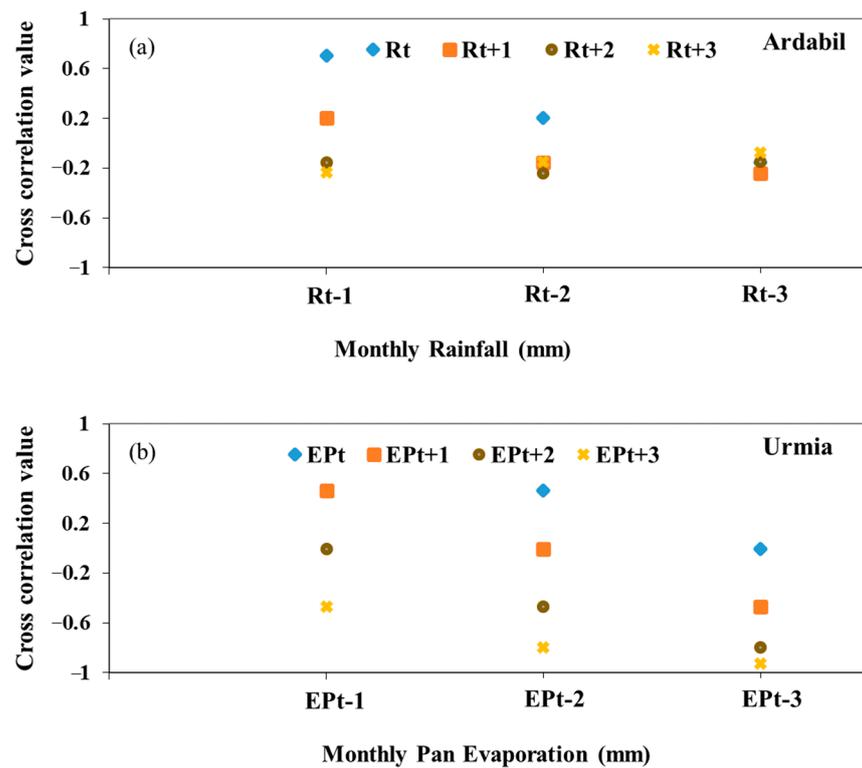


Figure 3. Cross-correlation analysis plots between input and output variables at (a) Ardabil and (b) Urmia stations.

### 2.2. Support Vector Machine (SVM)

Support vector machines (SVMs) were developed in 1992 by the Russian mathematician Vapnik, based on statistical learning theory. SVM is one of the learning methods with supervised learning which is used to analyze data that are implemented for regression analysis and classification. This learning system is used to both classify and predict the data fitness function to minimize errors in the data classification or fitness function. In linear data classification, an attempt is made to select a line that has a more reliable margin. Support vectors are the closest training points at the edge of the cloud and are used to define the boundary between classes [24]. If the data are linear and separate, the SVM uses linear machines to separate and train an optimal level with the least error and the maximum distance between the page and the nearest training points (support vectors) [24]. The structure of a support vector machine is shown in Figure 4 [25].

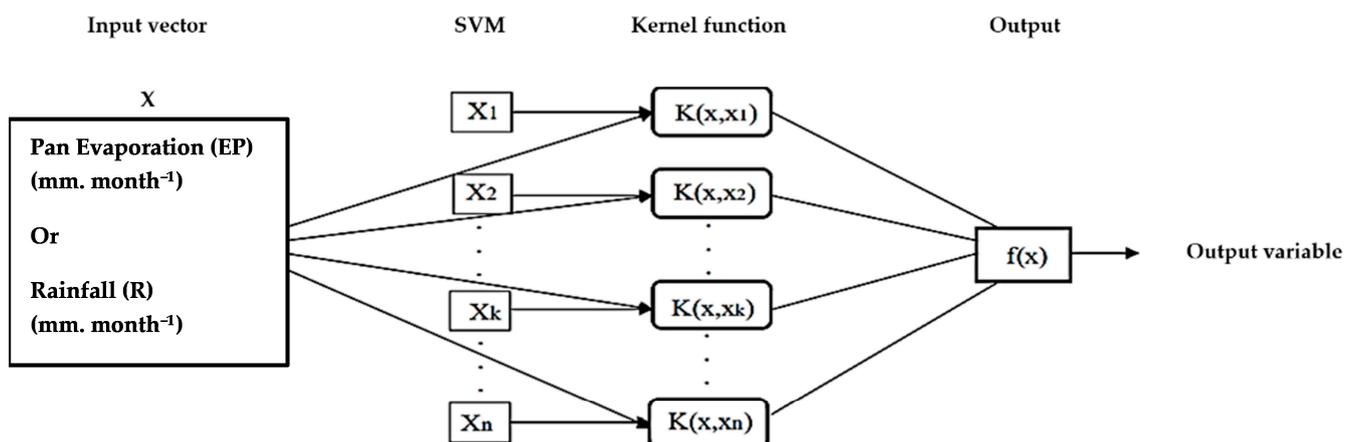


Figure 4. Support vector machine structure.

If the training points are in the form of  $[x_i, y_i]$ , and the input vector is in the form of  $x \in R^n$ , then the value of each class is defined as  $y_i \in [-1, 1] \ i = 1, \dots, i$ . The decision rules that can then be expressed by an optimal page that separates binary decision classes can be expressed as Equation (1):

$$Y = \text{sgn}(\sum_{i=1}^N y_i a_i (X \times X_i) + b) \quad (1)$$

In the above relation,  $Y$  is the output of the relation,  $y_i$  is the value of the sample class and  $X_i$ ,  $a_i$  and  $b$  are the parameters that determine the hyperplane. If linear separation is not possible, then Equation (1) is changed as follows:

$$Y = \text{sgn}(\sum_{i=1}^N y_i a_i K(X \times X_i) + b) \quad (2)$$

In Equation (2),  $K(X \times X_i)$  is a kernel function that generates internal multiplications to create SVM models with different modes of nonlinear decision levels in the data space, and for this purpose, it is necessary to define the line equation. The line equation in the 2D space is calculated by Equation (3), the plane equation is calculated by Equation (4) and the screen equation is calculated by Equation (5) [26].

$$w_1 x_1 + w_2 x_2 + b = 0 \quad (3)$$

$$w_1 x_1 + w_2 x_2 + w_3 x_3 + b = 0 \quad (4)$$

$$\sum_{i=1}^n w_i x_i + b = 0 \rightarrow w^T x + b = 0, w = \begin{bmatrix} w_1 \\ \vdots \\ w_i \end{bmatrix}, x = \begin{bmatrix} x_1 \\ \vdots \\ x_i \end{bmatrix} \quad (5)$$

According to Figure 5, the continuous bold line with the equation  $w^T x + b = 0$  is known as the line separating the data on the plane and divides them into two categories, A and B. This line leads to the formation of a space in which the data belonging to category A take a positive number and the data belonging to category B take a negative number. However, in SVM models, in addition to using the delimiter line, a confidence margin is also used for classification (Figure 5). In this case, none of the data are allowed to be in the middle area. Assuming that the line with the equation  $w^T x + b = 0$  is a boundary zero point, then for the data, depending on the position in classes A and B, respectively, the equations  $w^T x + b > 1$  and  $w^T x + b < -1$  are established. The thickness of the separator in the SVM includes an area and makes the classification process more resistant to the risk of misalignment [27].

One of the common methods for solving nonlinear problems is to use kernel functions. In fact, with a nonlinear transformation of the input space into a larger space, usury issues can be separated linearly. The choice of the kernel function is very important in SVM models, and different issues can be considered depending on the nature of the problem. Therefore, a function cannot be definitively introduced as the most suitable function for SVM. The types of important kernel functions that are common in engineering applications [28–30] and were used in the present study are shown in Table 2.

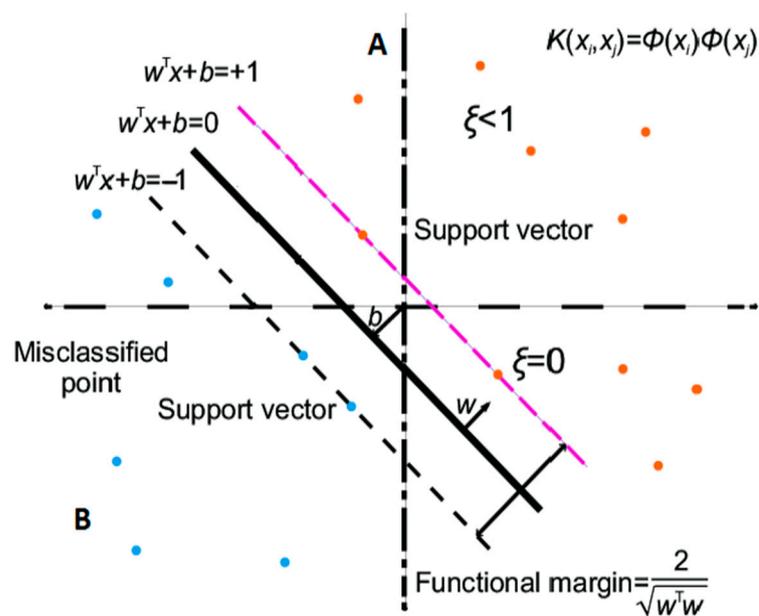


Figure 5. Graphic display of support vectors.

Table 2. The kernel functions used in the present study.

| Function Type                      | Kernel Function   |
|------------------------------------|---|
| Polynomial kernel function         | $k(x_i, x_j) = (x_i^T x_j + 1)^p$   |
| Radial-based kernel function (RBF) | $k(x_i, x_j) = e^{-\gamma \ x_i - x_j\ ^2}$   |
| Pearson kernel function (PUK)      | $k(x_i, x_j) = \frac{1}{[1 + (2\sqrt{ x_i - x_j ^2} \times \sqrt{2\frac{1}{\sigma}})]^2}$ |

### 2.3. Model Performance Evaluation Indicators

To evaluate the accuracy of SVM model scenarios according to Equations (6)–(8), the performance evaluation criteria of root mean square error (RMSE), coefficient of determination (CC) and Kling–Gupta (KGE) were used. The RMSE measures the best fit with the priority of high values of monthly rainfall and EP [31,32]. The KGE criterion in Equation (8) is one of the new criteria in the evaluation of hydrological models proposed by [33] and is in fact a modified version of the Nash–Sutcliffe efficiency (NSE) index.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

$$CC = \frac{(\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i)}{(\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2) (\sum_{i=1}^n y_i^2 - \frac{1}{n} (\sum_{i=1}^n y_i)^2)} \quad (7)$$

$$KGE = 1 - \sqrt{(cc - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (8)$$

In Equations (6)–(8),  $x_i$  and  $y_i$  are real and estimated values,  $n$  is the number of data evaluated, CC is the linear correlation coefficient between  $x_i$  and  $y_i$ ,  $\alpha$  is equal to the ratio of standard deviation  $y_i$  to standard deviation  $x_i$  and  $\beta$  is equal to the ratio of average  $y_i$  to average  $x_i$ .

### 3. Results and Discussion

#### 3.1. Determining the Most Effective Input Compounds to Model Output Aims

In order to determine the input of the SVM model, with the aim of predicting the current and lead times of one–three monthly rainfall (R) and pan evaporation (EP) events, the cross-correlations of lag times ( $R_{t-1}$ ,  $R_{t-2}$ ,  $R_{t-3}$  or  $EP_{t-1}$ ,  $EP_{t-2}$ ,  $EP_{t-3}$ ) against the target months ( $R_t$ ,  $R_{t+1}$ ,  $R_{t+2}$ ,  $R_{t+3}$  or  $EP_t$ ,  $EP_{t+1}$ ,  $EP_{t+2}$ ,  $EP_{t+3}$ ) were examined, and the results are shown in Figure 3. Therefore, for rainfall, various combinations of lag times to predict the current and lead times of one–three monthly rainfall (R) events were considered as the input of the SVM model and are listed in Tables 3 and 4. For each of the predicted time steps, we have a total of six scenarios. Of these six scenarios, three scenarios are shown in Table 3, and three scenarios are shown in Table 4. The scenarios in Table 3 show the different combinations of rainfall (R) lag times as input to the SVM model. According to the scenarios in Table 4, in addition to lag times, the month number ( $\tau$ ) was placed as an input next to other inputs to check the results of the scenarios without and with the month number. This is also true for the prediction of pan evaporation (EP), which can be seen in Tables 5 and 6.

**Table 3.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of monthly rainfall prediction in scenarios without the month number ( $\tau$ ) during training and testing periods at Ardabil station.

| Input Scenario                          | Output    | Model | Training     |              |                                 | Testing      |              |                                 |
|---|-----------|-------|--------------|--------------|---------------------------------|--------------|--------------|---------------------------------|
|   |           |       | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) |
| (1) $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$   | $R_t$     | SVM-1 | 0.863        | 0.820        | 6.832                           | 0.804        | 0.799        | 6.464                           |
| * (3) $R_{t-1}$ , $R_{t-2}$             | $R_t$     | SVM-3 | <b>0.839</b> | <b>0.779</b> | <b>7.391</b>                    | <b>0.815</b> | <b>0.807</b> | <b>6.230</b>                    |
| (5) $R_{t-1}$                           | $R_t$     | SVM-5 | 0.717        | 0.609        | 9.459                           | 0.666        | 0.615        | 7.912                           |
| (1) $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$   | $R_{t+1}$ | SVM-1 | 0.580        | 0.417        | 11.061                          | 0.454        | 0.377        | 9.694                           |
| * (3) $R_{t-1}$ , $R_{t-2}$             | $R_{t+1}$ | SVM-3 | <b>0.502</b> | <b>0.309</b> | <b>11.806</b>                   | <b>0.438</b> | <b>0.306</b> | <b>9.560</b>                    |
| (5) $R_{t-1}$                           | $R_{t+1}$ | SVM-5 | 0.240        | -0.098       | 13.177                          | 0.083        | -0.203       | 10.538                          |
| * (1) $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$ | $R_{t+2}$ | SVM-1 | <b>0.373</b> | <b>0.120</b> | <b>12.552</b>                   | <b>0.291</b> | <b>0.065</b> | <b>10.067</b>                   |
| (3) $R_{t-1}$ , $R_{t-2}$               | $R_{t+2}$ | SVM-3 | 0.277        | 0.025        | 13.190                          | 0.274        | 0.057        | 10.120                          |
| (5) $R_{t-1}$                           | $R_{t+2}$ | SVM-5 | 0.152        | -0.171       | 13.417                          | 0.185        | -0.151       | 10.255                          |
| * (1) $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$ | $R_{t+3}$ | SVM-1 | <b>0.411</b> | <b>0.154</b> | <b>12.430</b>                   | <b>0.401</b> | <b>0.142</b> | <b>9.603</b>                    |
| (3) $R_{t-1}$ , $R_{t-2}$               | $R_{t+3}$ | SVM-3 | 0.315        | 0.035        | 12.893                          | 0.345        | 0.066        | 9.847                           |
| (5) $R_{t-1}$                           | $R_{t+3}$ | SVM-5 | 0.265        | 0.007        | 13.252                          | 0.345        | 0.093        | 9.840                           |

\* The results in bold show the selected model.

**Table 4.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of monthly rainfall prediction in scenarios with the month number ( $\tau$ ) during training and testing periods at Ardabil station.

| Input Scenario                                   | Output    | Model | Training     |              |                                 | Testing      |              |                                 |
|--|-----------|-------|--------------|--------------|---------------------------------|--------------|--------------|---------------------------------|
|  |           |       | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) |
| * (2) $\tau$ , $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$ | $R_t$     | SVM-2 | <b>0.911</b> | <b>0.841</b> | <b>5.661</b>                    | <b>0.846</b> | <b>0.845</b> | <b>5.815</b>                    |
| (4) $\tau$ , $R_{t-1}$ , $R_{t-2}$               | $R_t$     | SVM-4 | 0.902        | 0.850        | 5.864                           | 0.832        | 0.826        | 6.199                           |
| (6) $\tau$ , $R_{t-1}$                           | $R_t$     | SVM-6 | 0.860        | 0.775        | 6.948                           | 0.812        | 0.811        | 6.428                           |
| * (2) $\tau$ , $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$ | $R_{t+1}$ | SVM-2 | <b>0.771</b> | <b>0.665</b> | <b>8.602</b>                    | <b>0.704</b> | <b>0.703</b> | <b>7.974</b>                    |
| (4) $\tau$ , $R_{t-1}$ , $R_{t-2}$               | $R_{t+1}$ | SVM-4 | 0.744        | 0.639        | 9.045                           | 0.709        | 0.709        | 8.013                           |
| (6) $\tau$ , $R_{t-1}$                           | $R_{t+1}$ | SVM-6 | 0.697        | 0.557        | 9.790                           | 0.650        | 0.633        | 8.403                           |
| * (2) $\tau$ , $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$ | $R_{t+2}$ | SVM-2 | <b>0.706</b> | <b>0.551</b> | <b>9.602</b>                    | <b>0.656</b> | <b>0.623</b> | <b>8.148</b>                    |
| (4) $\tau$ , $R_{t-1}$ , $R_{t-2}$               | $R_{t+2}$ | SVM-4 | 0.668        | 0.511        | 10.111                          | 0.653        | 0.628        | 8.244                           |
| (6) $\tau$ , $R_{t-1}$                           | $R_{t+2}$ | SVM-6 | 0.641        | 0.483        | 10.455                          | 0.632        | 0.604        | 8.467                           |
| (2) $\tau$ , $R_{t-1}$ , $R_{t-2}$ , $R_{t-3}$   | $R_{t+3}$ | SVM-2 | 0.723        | 0.579        | 9.356                           | 0.616        | 0.601        | 8.739                           |
| (4) $\tau$ , $R_{t-1}$ , $R_{t-2}$               | $R_{t+3}$ | SVM-4 | 0.684        | 0.532        | 9.860                           | 0.616        | 0.586        | 8.597                           |
| * (6) $\tau$ , $R_{t-1}$                         | $R_{t+3}$ | SVM-6 | <b>0.651</b> | <b>0.489</b> | <b>10.288</b>                   | <b>0.628</b> | <b>0.587</b> | <b>8.411</b>                    |

\* The results in bold show the selected model.

**Table 5.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of monthly EP prediction in scenarios without  $\tau$  during training and testing periods at Urmia station.

| Input Scenario  | Output            | Model | Training     |              |                                 | Testing      |              |                                 |
|---|-------------------|-------|--------------|--------------|---------------------------------|--------------|--------------|---------------------------------|
|   |                   |       | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) |
| * (1) EP <sub>t-1</sub> , EP <sub>t-2</sub> , EP <sub>t-3</sub> | EP <sub>t</sub>   | SVM-1 | <b>0.937</b> | <b>0.902</b> | <b>1.230</b>                    | <b>0.889</b> | <b>0.844</b> | <b>1.666</b>                    |
| (3) EP <sub>t-1</sub> , EP <sub>t-2</sub>                       | EP <sub>t</sub>   | SVM-3 | 0.901        | 0.864        | 1.543                           | 0.868        | 0.831        | 1.804                           |
| (5) EP <sub>t-1</sub>   | EP <sub>t</sub>   | SVM-5 | 0.828        | 0.818        | 1.969                           | 0.796        | 0.781        | 2.175                           |
| * (1) EP <sub>t-1</sub> , EP <sub>t-2</sub> , EP <sub>t-3</sub> | EP <sub>t+1</sub> | SVM-1 | <b>0.969</b> | <b>0.955</b> | <b>0.833</b>                    | <b>0.866</b> | <b>0.845</b> | <b>1.751</b>                    |
| (3) EP <sub>t-1</sub> , EP <sub>t-2</sub>                       | EP <sub>t+1</sub> | SVM-3 | 0.876        | 0.802        | 1.629                           | 0.807        | 0.761        | 1.990                           |
| (5) EP <sub>t-1</sub>   | EP <sub>t+1</sub> | SVM-5 | 0.564        | 0.494        | 2.842                           | 0.532        | 0.459        | 2.921                           |
| * (1) EP <sub>t-1</sub> , EP <sub>t-2</sub> , EP <sub>t-3</sub> | EP <sub>t+2</sub> | SVM-1 | <b>0.965</b> | <b>0.946</b> | <b>0.883</b>                    | <b>0.888</b> | <b>0.850</b> | <b>1.603</b>                    |
| (3) EP <sub>t-1</sub> , EP <sub>t-2</sub>                       | EP <sub>t+2</sub> | SVM-3 | 0.911        | 0.861        | 1.394                           | 0.797        | 0.764        | 2.072                           |
| (5) EP <sub>t-1</sub>   | EP <sub>t+2</sub> | SVM-5 | 0.421        | 0.266        | 3.106                           | 0.400        | 0.250        | 3.093                           |
| * (1) EP <sub>t-1</sub> , EP <sub>t-2</sub> , EP <sub>t-3</sub> | EP <sub>t+3</sub> | SVM-1 | <b>0.978</b> | <b>0.956</b> | <b>0.701</b>                    | <b>0.930</b> | <b>0.846</b> | <b>1.339</b>                    |
| (3) EP <sub>t-1</sub> , EP <sub>t-2</sub>                       | EP <sub>t+3</sub> | SVM-3 | 0.945        | 0.926        | 1.102                           | 0.869        | 0.834        | 1.712                           |
| (5) EP <sub>t-1</sub>   | EP <sub>t+3</sub> | SVM-5 | 0.689        | 0.536        | 2.439                           | 0.617        | 0.478        | 2.672                           |

\* The results in bold show the selected model.

**Table 6.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of monthly EP prediction in scenarios with  $\tau$  during training and testing periods at Urmia station.

| Input Scenario   | Output            | Model | Training     |              |                                 | Testing      |              |                                 |
|--|-------------------|-------|--------------|--------------|---------------------------------|--------------|--------------|---------------------------------|
|  |                   |       | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) |
| (2) $\tau$ , EP <sub>t-1</sub> , EP <sub>t-2</sub> , EP <sub>t-3</sub> | EP <sub>t</sub>   | SVM-2 | 0.984        | 0.969        | 0.602                           | 0.940        | 0.913        | 1.205                           |
| (4) $\tau$ , EP <sub>t-1</sub> , EP <sub>t-2</sub>                     | EP <sub>t</sub>   | SVM-4 | 0.979        | 0.970        | 0.678                           | 0.951        | 0.946        | 1.065                           |
| * (6) $\tau$ , EP <sub>t-1</sub>                                       | EP <sub>t</sub>   | SVM-6 | <b>0.976</b> | <b>0.955</b> | <b>0.731</b>                    | <b>0.957</b> | <b>0.938</b> | <b>0.996</b>                    |
| (2) $\tau$ , EP <sub>t-1</sub> , EP <sub>t-2</sub> , EP <sub>t-3</sub> | EP <sub>t+1</sub> | SVM-2 | 0.982        | 0.976        | 0.641                           | 0.920        | 0.899        | 1.370                           |
| (4) $\tau$ , EP <sub>t-1</sub> , EP <sub>t-2</sub>                     | EP <sub>t+1</sub> | SVM-4 | 0.980        | 0.968        | 0.666                           | 0.938        | 0.924        | 1.197                           |
| * (6) $\tau$ , EP <sub>t-1</sub>                                       | EP <sub>t+1</sub> | SVM-6 | <b>0.977</b> | <b>0.966</b> | <b>0.723</b>                    | <b>0.942</b> | <b>0.936</b> | <b>1.144</b>                    |
| (2) $\tau$ , R <sub>t-1</sub> , R <sub>t-2</sub> , R <sub>t-3</sub>    | EP <sub>t+2</sub> | SVM-2 | 0.982        | 0.975        | 0.649                           | 0.934        | 0.893        | 1.244                           |
| * (4) $\tau$ , EP <sub>t-1</sub> , EP <sub>t-2</sub>                   | EP <sub>t+2</sub> | SVM-4 | <b>0.980</b> | <b>0.974</b> | <b>0.676</b>                    | <b>0.944</b> | <b>0.919</b> | <b>1.144</b>                    |
| (6) $\tau$ , EP <sub>t-1</sub>   | EP <sub>t+2</sub> | SVM-6 | 0.977        | 0.959        | 0.722                           | 0.942        | 0.912        | 1.173                           |
| (2) $\tau$ , EP <sub>t-1</sub> , EP <sub>t-2</sub> , EP <sub>t-3</sub> | EP <sub>t+3</sub> | SVM-2 | 0.982        | 0.978        | 0.644                           | 0.945        | 0.881        | 1.167                           |
| (4) $\tau$ , EP <sub>t-1</sub> , EP <sub>t-2</sub>                     | EP <sub>t+3</sub> | SVM-4 | 0.980        | 0.972        | 0.675                           | 0.949        | 0.902        | 1.102                           |
| * (6) $\tau$ , EP <sub>t-1</sub>                                       | EP <sub>t+3</sub> | SVM-6 | <b>0.979</b> | <b>0.968</b> | <b>0.684</b>                    | <b>0.951</b> | <b>0.918</b> | <b>1.076</b>                    |

\* The results in bold show the selected model.

### 3.2. Simulation of Monthly Rainfall (R) at Ardabil Station

In order to evaluate the strength of the developed support vector machine (SVM) model in predicting the monthly rainfall (R) of Ardabil station for the current and lead times of one–three monthly rainfall events, and in case the month number ( $\tau$ ) is not among the input scenarios, statistical indicators RMSE (mm. Month<sup>-1</sup>), KGE and CC were calculated for the test phase. The results of these indicators are listed in Table 3. According to Table 3, and in the absence of the month number ( $\tau$ ) as input in the scenarios, scenario 3 (SVM-3), with the input of two monthly rainfall lag times (R<sub>t-1</sub>, R<sub>t-2</sub>), was able to produce the least errors compared to other scenarios in predicting the current month's rainfall (R<sub>t</sub>), with RMSE = 6.23 mm. month<sup>-1</sup>, KGE = 0.807 and CC = 0.815. In predicting 1 month of upcoming rainfall (R<sub>t+1</sub>), again, scenario 3 (SVM-3) managed the lowest value of RMSE = 9.56 mm. month<sup>-1</sup> and had the highest values of KGE = 0.306 and CC = 0.438, compared to other scenarios in this time step.

In predicting the rainfall of the upcoming two months ( $R_{t+2}$ ), better results from scenario 1 (SVM-1) compared to the other scenarios, with  $RMSE = 10.067 \text{ mm. month}^{-1}$ ,  $KGE = 0.065$  and  $CC = 0.291$ , were observed. The last goal in Table 3 is to forecast the upcoming three months of rainfall ( $R_{t+3}$ ). According to the results of Table 3, it can be said that by targeting the upcoming one, two and three months of rainfall, the accuracy of the forecast with the same fixed scenarios was less than the target of the current month's rainfall ( $R_t$ ), from  $CC = 0.815$  in the current step ( $R_t$ ) to  $CC = 0.438$ ,  $0.291$  and  $0.401$  in the time steps, respectively, reached for  $R_{t+1}$ ,  $R_{t+2}$  and  $R_{t+3}$ . Another point related to the results of Table 3 is that the selected scenario is the same in the current time steps and upcoming month of rainfall ( $R_t$ ,  $R_{t+1}$ ), and scenario 3 (SVM-3) in both steps is selected as the best scenario. This happened in the time steps of the upcoming two and three months ( $R_{t+2}$ ,  $R_{t+3}$ ) for scenario 1 (SVM-1).

In the next step, considering the month number ( $\tau$ ) as input within scenarios 1, 3 and 5 (SVM-1, SVM-3, SVM-5), new scenarios 2, 4 and 6 (SVM-2, SVM-4, SVM-6) were obtained. New scenarios were introduced as input to the SVM model. The results of the statistical indicators obtained from the new scenarios are shown in Table 4. According to Table 4, it is generally observed that the results in all time steps significantly improved compared to the case where the month number ( $\tau$ ) was not among the inputs. In Table 4, and the current time step ( $R_t$ ), scenario 2 (SVM-2), with inputs ( $\tau$ ,  $R_{t-1}$ ,  $R_{t-2}$ ,  $R_{t-3}$ ) and the lowest  $RMSE = 5.815 \text{ mm. month}^{-1}$ , showed the highest values of  $KGE = 0.845$  and  $CC = 0.846$  compared to scenarios 4 and 6 (SVM-4 and SVM-6); compared to the result of the superior scenario 3 (SVM-3) related to the current time step ( $R_t$ ) in Table 3, we saw an improvement of 6.7, 4.7 and 3.8% in the  $RMSE$ ,  $KGE$  and  $CC$  statistical indices, respectively, in the case that the month number ( $\tau$ ) was considered as input in the scenarios. In the next time step, i.e., rainfall for the upcoming month ( $R_{t+1}$ ), scenario 2 (SVM-2), with  $RMSE = 7.974 \text{ mm. month}^{-1}$ ,  $KGE = 0.703$  and  $CC = 0.704$ , was selected as the best scenario in this time step, and the  $RMSE$ ,  $KGE$  and  $CC$  indices improved by 16.6, 129 and 60.7%, respectively, compared to the best scenario 3 (SVM-3) for the month number ( $\tau$ ). According to Table 4, scenario 2 (SVM-2), as the best scenario, with the lowest  $RMSE = 8.148 \text{ mm. month}^{-1}$  and the highest values of  $KGE = 0.623$  and  $CC = 0.656$ , compared to the other scenarios, was selected to predict rainfall for the upcoming two months ( $R_{t+2}$ ) at Ardabil station. In this time step, we also saw a 19.1, 858 and 125% improvement in  $RMSE$ ,  $KGE$  and  $CC$  values, respectively, compared to the best scenario 1 (SVM-1) in the case without  $\tau$  in the same time step (Table 3). In the last target time step, i.e., the upcoming three months of rainfall time step ( $R_{t+3}$ ), scenario 6 (SVM-6), with  $RMSE = 8.411 \text{ mm. month}^{-1}$ ,  $KGE = 0.587$  and  $CC = 0.628$ , was identified as the best scenario, and we saw an improvement of 12.4, 313 and 56.6% in  $RMSE$ ,  $KGE$  and  $CC$  statistical indices, respectively, compared to the best scenario in Table 3.

In order to carefully examine the increase in the accuracy of the SVM model in predicting the time steps of the upcoming one, two and three months, in the presence and absence of the month number ( $\tau$ ) input, time series, observational scatter plots and rainfall forecasts for the top scenarios with the month are drawn in Table 4, and the corresponding scenarios without the month number ( $\tau$ ) in Table 3 are drawn according to Figure 6 for the test period. According to Figure 6, and to the observed time series diagrams and rainfall forecast at the current time step ( $R_t$ ), it is observed that the forecasted data in the scenario mode with the number of the month (SVM-2) are more consistent with the observational data. The time step scatter plot diagram also confirms this. Therefore, it is observed that the regression line related to the scenario with the month number (SVM-2) is closer to the bisector line than the scenario without the month number (SVM-1), and less overestimation is observed in this scenario. The improvement in the results in the scenario with the month ( $\tau$ ) in rainfall prediction with a time step of one month later ( $R_{t+1}$ ) is very evident in the observational and forecast time series diagram; compared to the no-month scenario (SVM-1), there is more compliance with the observational data. In the observed and predicted scatter diagrams of one upcoming month ( $R_{t+1}$ ), the scenario regression line

with the month (SVM-2) is much closer to the bisector line than the scenario without the month (SVM-1). Additionally, it requires fewer estimation changes.

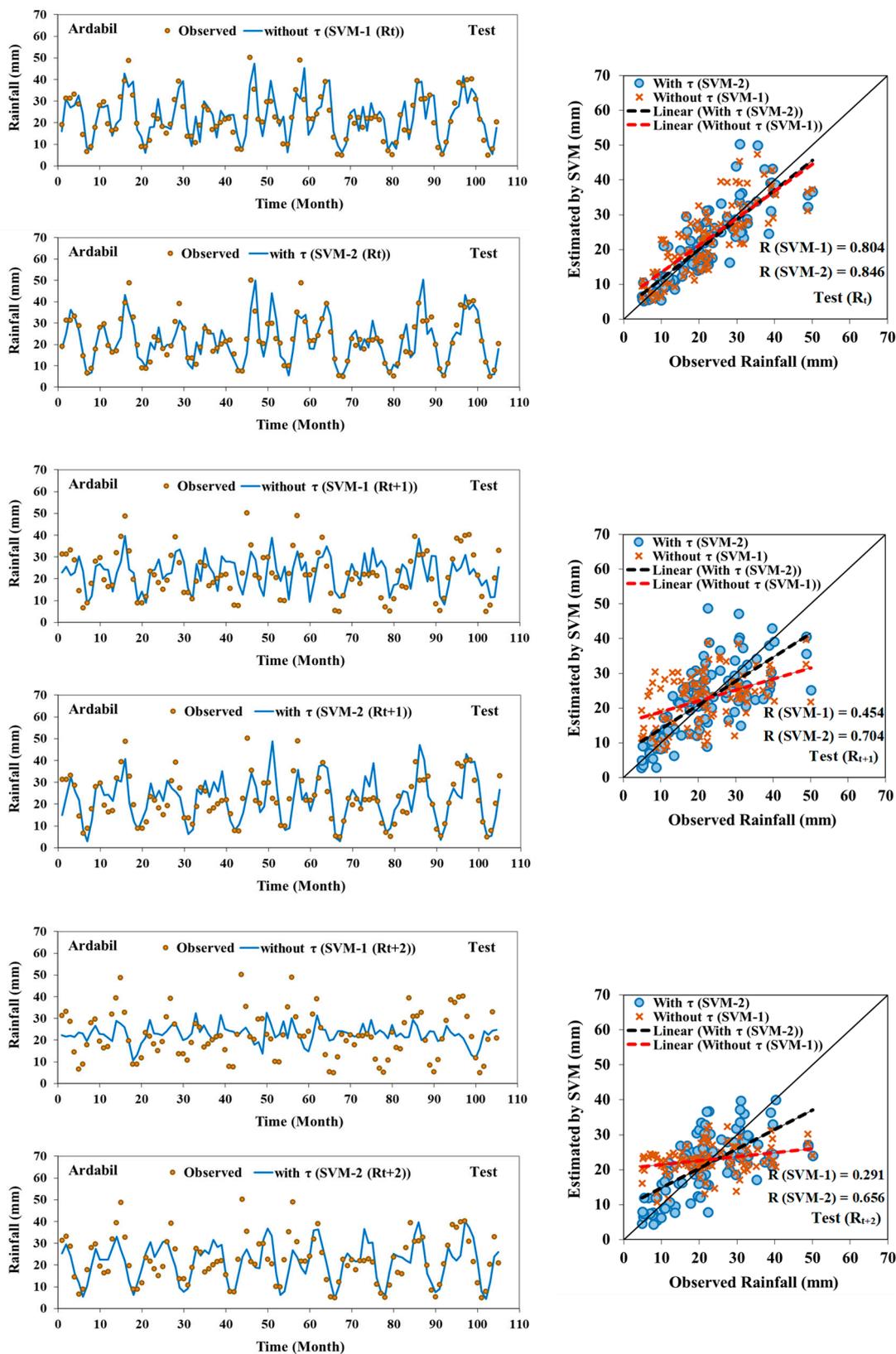
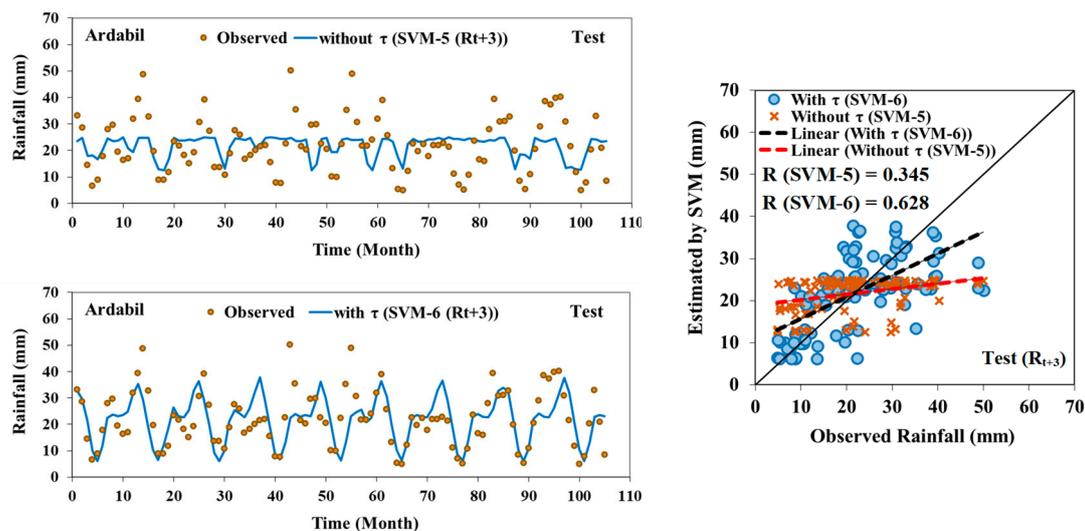


Figure 6. Cont.



**Figure 6.** Time series and scatter plots between observed and simulated current and lead times of 1–3 monthly rainfall values by SVM best scenarios without and with  $\tau$  input (Tables 3 and 4) during testing period at Ardabil station.

As the rainfall forecast time steps at Ardabil station move forward, the accuracy of scenarios with the same month as the results of the statistical indicators becomes more visible. As it is shown in the rainfall data distribution diagram for the upcoming two months, and as predicted by the scenario without the month number (SVM-2) and with the month number (SVM-1), the improvement in the results by the SVM-2 model is very obvious, and the regression line of this scenario is closer to the bisector line than SVM-1; therefore, the accuracy of estimating rainfall for the upcoming two months ( $R_{t+2}$ ) is higher. In Figure 6, the very poor results of forecasting the upcoming three months of rainfall ( $R_{t+3}$ ) by the scenario without the month number (SVM-5) in the time series diagrams, as well as the observed and predicted distribution, are clear, and the scenario with the month number (SVM-6) could improve the results better than the SVM-5 scenario and increase the accuracy of the forecast. Finally, it can be said that adding a simple characteristic such as the month number ( $\tau$ ) in the delay scenarios, with the aim of predicting the rainfall of Ardabil station in different time steps, enabled dramatically improving the results, especially in the step of the upcoming one, two and three months.

### 3.3. Simulation of Monthly Pan Evaporation (EP) at Urmia Station

In order to investigate the changes in the case of a simple characteristic such as the month number ( $\tau$ ) in the time delay input scenarios for the monthly estimation of another parameter except rainfall, the SVM model was implemented for the monthly pan evaporation (EP) of Urmia station. Initially, in order to predict the evaporation from the EP of Urmia station for four time steps (current month, and upcoming one, two and three months), the SVM model was implemented for scenarios that do not have the month number ( $\tau$ ), and the results of the statistical indicators obtained from this work are listed in Table 5. According to Table 5, in estimating the evaporation from the current lunar pan ( $EP_t$ ), scenario 1, with inputs of one-, two- and three-month lag times of EP ( $EP_{t-1}$ ,  $EP_{t-2}$ ,  $EP_{t-3}$ ), had the lowest RMSE = 1.666 mm. month<sup>-1</sup> and the highest values of KGE = 0.844 and CC = 0.889, compared to other scenarios, and it should be selected as the best scenario in this time step. In predicting EP with the aim of predicting the time step of the upcoming month ( $EP_{t+1}$ ), again, scenario 1 (SVM-1), with RMSE = 1.751 mm. month<sup>-1</sup>, KGE = 0.845 and CC = 0.866, showed better accuracy in this time step. In a step ahead of time, i.e., predicting EP for the upcoming two months ( $EP_{t+2}$ ), scenario 1 (SVM-1), due to having the lowest value of RMSE = 1.603 mm. month<sup>-1</sup> and the highest values of KGE = 0.85 and CC = 0.888, appeared more successful than the other scenarios in this time

step. In the last step, for the prediction of EP for the upcoming three months ( $EP_{t+3}$ ), again, scenario 1 (SVM-1), as the best scenario, with  $RMSE = 1.339 \text{ mm. month}^{-1}$ ,  $KGE = 0.846$  and  $CC = 0.93$ , was selected. As a result, based on the results of Table 5, it is observed that unlike rainfall, in predicting EP with increasing time steps, the accuracy of the SVM model did not decrease much, and the model has a good accuracy. Even the upcoming three months' time step of EP ( $EP_{t+3}$ ) was able to achieve the best result in the same scenario 1 (SVM-1) compared to the current month's time step ( $EP_t$ ), and this is interesting. Additionally, scenario 1 (SVM-1), with inputs of one-, two- and three-month lag times of EP ( $EP_{t-1}$ ,  $EP_{t-2}$ ,  $EP_{t-3}$ ), could be selected as the best scenario in all time steps; other scenarios that were examined were not better than that.

The results in Table 5 show the high power of the SVM model in the monthly forecast of the EP process at Urmia station under scenarios 1, 3 and 5. However, in order to evaluate the possibility of increasing the accuracy in predicting this parameter, the month number ( $\tau$ ) was considered as input in the mentioned scenarios, and the accuracy of the scenarios was measured again; the obtained results of the statistical indicators are presented in Table 6. The results of evaluating the statistical indicators in Table 6 show that the accuracy of the SVM model increases again in all scenarios when considering the month number ( $\tau$ ) along with other inputs. If the case is viewed similarly to the prediction of the monthly rainfall process at Ardabil station in different future time steps, in the top scenarios in Table 6, there is increased accuracy in predicting EP for three lead times. When the month number was applied to the scenarios with the aim of increasing the accuracy in predicting EP in the current month ( $EP_t$ ), the best scenario, scenario 6, with the input of the month number and one-month delayed evaporation from the monthly pan ( $\tau$ ,  $EP_{t-1}$ ), with  $RMSE = 0.996 \text{ mm. month}^{-1}$ ,  $KGE = 0.938$  and  $CC = 0.957$ , was determined.

Compared to the best result set out in Table 4 for this time step, the SVM model (SVM-6) was able to use the month number to accurately predict the RMSE, KGE and CC of the current month's EP ( $EP_t$ ), seeing improvements of 40.2, 11.1 and 7.6%. In the time step of the upcoming month ( $EP_{t+1}$ ), again, scenario 6, with  $RMSE = 1.144 \text{ mm. month}^{-1}$ ,  $KGE = 0.936$  and  $CC = 0.942$ , returned the best results among the scenarios in this time step; compared to the best scenario in Table 5, the accuracy of RMSE, KGE and CC indices increased by 34.7, 10.8 and 8.8%, respectively. In the next modeling step, which was performed with the aim of predicting EP for the next two months ( $EP_{t+2}$ ), scenario 4 (SVM-4), with the input of the month number and lag times of one and two months ( $\tau$ ,  $EP_{t-1}$ ,  $EP_{t-2}$ ), and with  $RMSE = 1.144 \text{ mm. month}^{-1}$ ,  $KGE = 0.919$  and  $CC = 0.944$ , was selected as the best scenario. Comparing the results of these indicators with the indicators of the best scenario for predicting the time step, we see 28.6, 8.1 and 6.3% improvements in the results if the number of the month ( $\tau$ ) is used in the input of the scenarios with time delays. In the last time step evaluated, which is the prediction of evaporation from the pan for the upcoming three months ( $EP_{t+3}$ ), scenario 6 (SVM-6), with  $RMSE = 1.076 \text{ mm. month}^{-1}$ ,  $KGE = 0.918$  and  $CC = 0.951$ , was selected as the best scenario; compared to the best result in the case without using  $\tau$ , increases of 19.6, 8.5 and 2.3% in the accuracy of the evaluation indicators, respectively, were found. Finally, it can be said that in predicting the monthly EP in the case that the month number was added to the lag time scenarios, not only do we not see a decrease in the forecasting accuracy but the accuracy increases to the extent that it persuades the user to use this simple and effective solution.

Evaluation of the scenarios in the test phase in the presence of the month number is also shown in Figure 7. Thus, the best scenarios selected from Table 6 compared to their unnumbered scenarios of the month were examined in terms of time series diagrams and the observational distribution and prediction in all time steps. As it is shown in Figure 7, if the monthly prediction of the current time step EP ( $EP_t$ ) is examined, the observed and predicted time series diagrams show a high correlation of the predicted data under the scenario containing the number. The scenario with the month number (SVM-6) used observational data and was able to detect the evaporation trend well compared to the scenario without  $\tau$  (SVM-5). The observational scatter plot and prediction of the current

month's EP ( $EP_t$ ) under the SVM-5 and SVM-6 scenarios show that the SVM-6 scenario regression line is more in line with the bisector line than the SVM-5 scenario. Estimation and underestimation in the scenario with  $\tau$  (SVM-6) are less than those of SVM-5. The increased accuracy in predicting EP under the scenario with  $\tau$  in the time step of the next month ( $EP_{t+1}$ ) compared to the scenario without  $\tau$ , as shown in Figure 7, in both time series diagrams and the observational distribution and forecast, is completely clear. The SVM-6 scenario was able to predict  $EP_{t+1}$  and establish high compliance with the observational data. The issue in the distribution chart is also clearly defined, and the SVM-6 scenario's regression line is much more compatible than that of SVM-5. Visual examination of time series diagrams and the observational distribution and prediction of the upcoming two months of EP ( $EP_{t+2}$ ) also indicates that the forecast accuracy is increased by the SVM-4 scenario compared to the SVM-3 scenario. In terms of the observational distribution and prediction diagrams, the scenario with  $\tau$  (SVM-4) was able to overcome the weakness of overestimation found in the SVM-3 scenario, to some extent. In the last time step in Figure 7, i.e., the upcoming three months of EP ( $EP_{t+3}$ ), the increase in the accuracy of the scenario with  $\tau$  (SVM-6) compared to the scenario without  $\tau$  (SVM-5), in both time series and distribution diagrams, is evident. According to the observational distribution chart and forecast for the upcoming three months ( $EP_{t+3}$ ), the SVM-6 scenario, which contains the month number, was able to eliminate the overestimation and underestimation of scenario 5 (SVM-5) without  $\tau$ , with high accuracy.

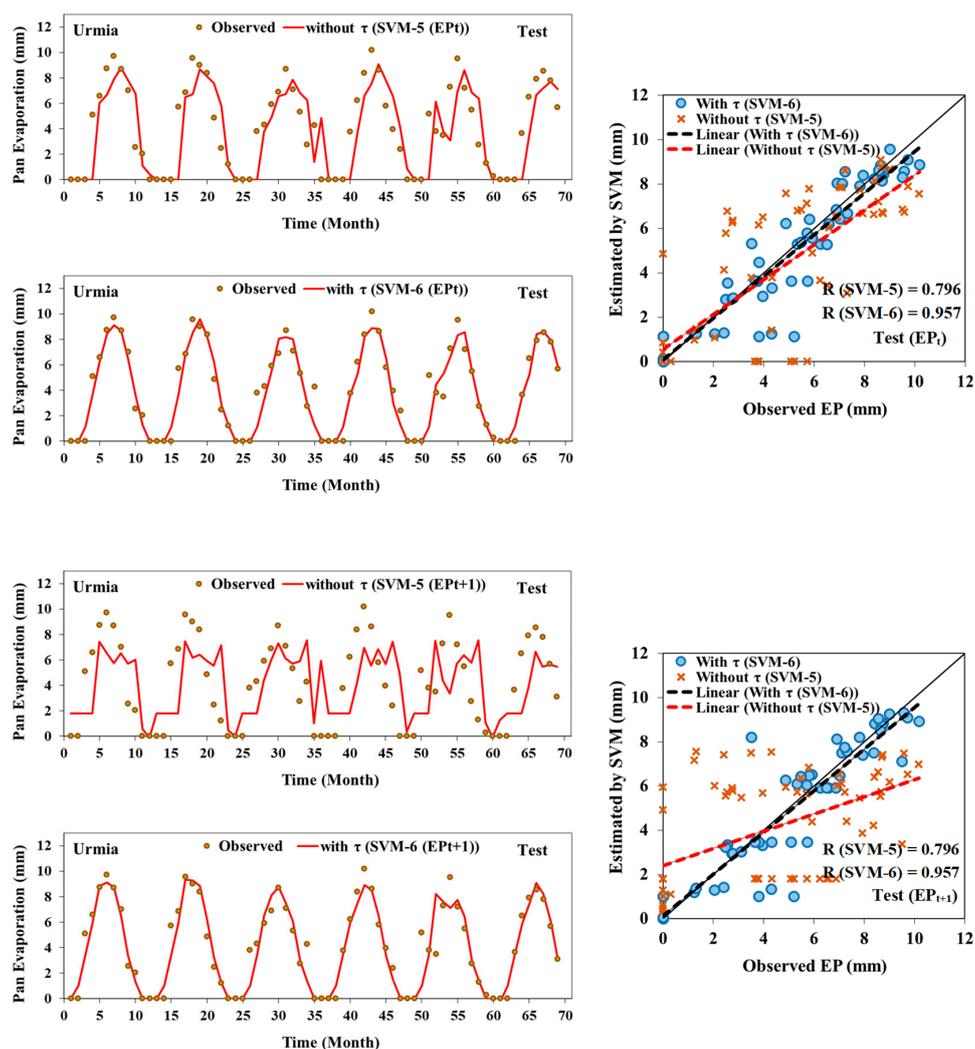
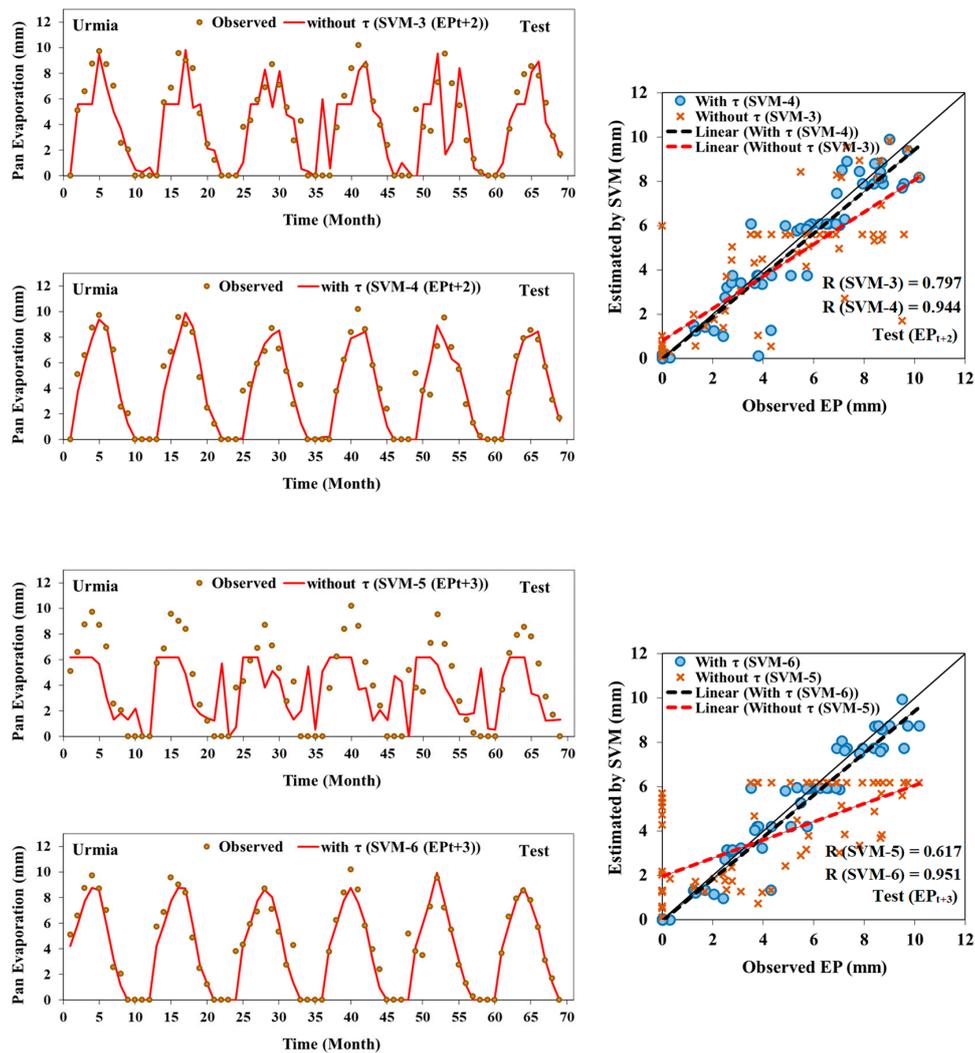


Figure 7. Cont.



**Figure 7.** Time series and scatter plots between observed and simulated current and lead times of 1–3 monthly EP values by SVM best scenarios without and with  $\tau$  input (Tables 5 and 6) during testing period at Urmia station.

### 3.4. Simulation for Months with Maximum Rainfall and Pan Evaporation

According to the results of the previous sections, it was observed that the month number ( $\tau$ ) characteristic was able to improve the prediction accuracy of both rainfall (R) and pan evaporation (EP) parameters under the SVM model when entering the scenarios. In order to describe the potential physical or numerical reasons for these results, a study was conducted on how to improve the accuracy of the SVM model, especially under specific situations. According to the observed precipitation and evaporation time series diagrams in Figure 2, a seasonal pattern is observed, especially for pan evaporation (EP). The reason for the increased accuracy of the SVM model in the case of entering the month number ( $\tau$ ) can be due to the strong seasonal signal in the time series. For example, according to Figure 2, there is a clear pattern whereby certain months of the year have extremely low monthly average values. As a result, it is doubtful that the SVM model, by adding the month number ( $\tau$ ), detects this numerical pattern and produces high-precision results. Given that the months that contain the highest values are physically one of the most important characteristics of hydrological variables, in the following, we will examine how the SVM model will perform in the presence or absence of the month number under these special conditions. For this purpose, first, the highest monthly values (peak points) of the rainfall (70 months) and pan evaporation (28 months) variables were extracted according to the time series (Figure 2) related to each. In order to compare the results with the results

of Sections 3.2 and 3.3, the input scenarios and target time steps are exactly the same as those mentioned. To enter the SVM model, the data were divided under these conditions, meaning that 60% of the data were separated for the training phase, and 40% for the testing phase.

In this regard, the results of the statistical indicators under these specific conditions with three target time steps, for rainfall (R) in Tables 7 and 8 (respectively, without and with  $\tau$ ), and for pan evaporation (EP) in Tables 9 and 10 (respectively, without and with  $\tau$ ), were collected. In general, it can be seen that the SVM model was unable to create high accuracy under these new special conditions, one of the reasons for which could be related to the small amount of data in the training phase. In order to evaluate the improvement or deterioration in the results when considering the month number, first, for the precipitation variable, according to Tables 7 and 8, it can be stated that by using the month number, the SVM model was able to improve the results for almost all targeted time steps (except  $R_{t+2}$ ). For the best conditions ( $R_t$ ), the accuracy was improved by 17% according to the RMSE index. Under these special conditions, and due to the more random rainfall time series, an acceptable score can be given to the SVM model. For the pan evaporation variable, according to Tables 9 and 10, although the results under the conditions of using the highest values are weaker than the application mode of all data, the results improved in the case of considering the month number compared to the cases without the month number (except  $EP_{t+3}$ ). For the best conditions ( $EP_{t+2}$ ), the accuracy was improved by 13% according to the RMSE index. According to the above results, it is observed that due to the fact that the precipitation variable is more random, using the highest values, the improvement in the results with the month number was greater than the evaporation variable. According to the observational pan evaporation time series in Figure 2, this time series is completely periodic, and the SVM model could provide ideal results by relying on this; however, in these special circumstances, the month number proved its ability to improve the results.

**Table 7.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of highest monthly rainfall values in scenarios without month number ( $\tau$ ) during training and testing periods at Ardabil station.

| Input Scenario                  | Output    | Model | Training     |               |                                    | Testing       |               |                                    |
|---------------------------------|-----------|-------|--------------|---------------|------------------------------------|---------------|---------------|------------------------------------|
|                                 |           |       | CC           | KGE           | RMSE<br>(mm. Month <sup>-1</sup> ) | CC            | KGE           | RMSE<br>(mm. Month <sup>-1</sup> ) |
| (1) $R_{t-1}, R_{t-2}, R_{t-3}$ | $R_t$     | SVM-1 | 0.759        | 0.626         | 7.873                              | 0.135         | 0.133         | 14.853                             |
| (3) $R_{t-1}, R_{t-2}$          | $R_t$     | SVM-3 | 0.614        | 0.540         | 9.663                              | 0.002         | 0.0001        | 15.733                             |
| * (5) $R_{t-1}$                 | $R_t$     | SVM-5 | <b>0.540</b> | <b>0.309</b>  | <b>10.280</b>                      | <b>-0.062</b> | <b>-0.159</b> | <b>13.378</b>                      |
| (1) $R_{t-1}, R_{t-2}, R_{t-3}$ | $R_{t+1}$ | SVM-1 | 0.718        | 0.467         | 8.822                              | -0.290        | -0.371        | 15.698                             |
| (3) $R_{t-1}, R_{t-2}$          | $R_{t+1}$ | SVM-3 | 0.371        | 0.052         | 11.503                             | -0.112        | -0.264        | 13.285                             |
| * (5) $R_{t-1}$                 | $R_{t+1}$ | SVM-5 | <b>0.179</b> | <b>-0.153</b> | <b>12.096</b>                      | <b>-0.182</b> | <b>-0.443</b> | <b>11.695</b>                      |
| (1) $R_{t-1}, R_{t-2}, R_{t-3}$ | $R_{t+2}$ | SVM-1 | 0.814        | 0.606         | 7.577                              | -0.353        | -0.392        | 17.562                             |
| (3) $R_{t-1}, R_{t-2}$          | $R_{t+2}$ | SVM-3 | 0.519        | 0.263         | 10.565                             | -0.207        | -0.298        | 14.177                             |
| * (5) $R_{t-1}$                 | $R_{t+2}$ | SVM-5 | <b>0.197</b> | <b>-0.154</b> | <b>11.984</b>                      | <b>-0.300</b> | <b>-0.527</b> | <b>11.241</b>                      |
| (1) $R_{t-1}, R_{t-2}, R_{t-3}$ | $R_{t+3}$ | SVM-1 | 0.815        | 0.641         | 7.030                              | -0.198        | -0.226        | 15.986                             |
| (3) $R_{t-1}, R_{t-2}$          | $R_{t+3}$ | SVM-3 | 0.501        | 0.294         | 10.229                             | -0.267        | -0.347        | 14.576                             |
| * (5) $R_{t-1}$                 | $R_{t+3}$ | SVM-5 | <b>0.108</b> | <b>-0.240</b> | <b>11.740</b>                      | <b>-0.378</b> | <b>-0.627</b> | <b>11.990</b>                      |

\* The results in bold show the selected model.

**Table 8.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of highest monthly rainfall values in scenarios with month number ( $\tau$ ) during training and testing periods at Ardabil station.

| Input Scenario                        | Output    | Model | Training     |               |                                 | Testing       |               |                                 |
|---------------------------------------|-----------|-------|--------------|---------------|---------------------------------|---------------|---------------|---------------------------------|
|                                       |           |       | CC           | KGE           | RMSE (mm. Month <sup>-1</sup> ) | CC            | KGE           | RMSE (mm. Month <sup>-1</sup> ) |
| (2) $\tau, R_{t-1}, R_{t-2}, R_{t-3}$ | $R_t$     | SVM-2 | 0.912        | 0.795         | 5.238                           | 0.309         | 0.300         | 12.294                          |
| (4) $\tau, R_{t-1}, R_{t-2}$          | $R_t$     | SVM-4 | 0.754        | 0.633         | 7.931                           | 0.256         | 0.254         | 13.864                          |
| * (6) $\tau, R_{t-1}$                 | $R_t$     | SVM-6 | <b>0.489</b> | <b>0.209</b>  | <b>10.603</b>                   | <b>0.160</b>  | <b>-0.082</b> | <b>11.394</b>                   |
| (2) $\tau, R_{t-1}, R_{t-2}, R_{t-3}$ | $R_{t+1}$ | SVM-2 | 0.827        | 0.695         | 6.999                           | -0.050        | -0.078        | 15.353                          |
| (4) $\tau, R_{t-1}, R_{t-2}$          | $R_{t+1}$ | SVM-4 | 0.380        | 0.157         | 11.441                          | -0.203        | -0.337        | 13.524                          |
| * (6) $\tau, R_{t-1}$                 | $R_{t+1}$ | SVM-6 | <b>0.293</b> | <b>-0.013</b> | <b>11.783</b>                   | <b>0.026</b>  | <b>-0.177</b> | <b>11.512</b>                   |
| (2) $\tau, R_{t-1}, R_{t-2}, R_{t-3}$ | $R_{t+2}$ | SVM-2 | 0.897        | 0.733         | 5.698                           | 0.009         | -0.099        | 13.898                          |
| (4) $\tau, R_{t-1}, R_{t-2}$          | $R_{t+2}$ | SVM-4 | 0.649        | 0.500         | 9.506                           | 0.045         | -0.007        | 14.704                          |
| * (6) $\tau, R_{t-1}$                 | $R_{t+2}$ | SVM-6 | <b>0.455</b> | <b>0.256</b>  | <b>10.929</b>                   | <b>-0.020</b> | <b>-0.084</b> | <b>12.816</b>                   |
| (2) $\tau, R_{t-1}, R_{t-2}, R_{t-3}$ | $R_{t+3}$ | SVM-2 | 0.950        | 0.827         | 3.939                           | 0.055         | -0.038        | 13.888                          |
| (4) $\tau, R_{t-1}, R_{t-2}$          | $R_{t+3}$ | SVM-4 | 0.752        | 0.526         | 8.020                           | 0.063         | -0.003        | 13.175                          |
| * (6) $\tau, R_{t-1}$                 | $R_{t+3}$ | SVM-6 | <b>0.484</b> | <b>0.243</b>  | <b>10.448</b>                   | <b>-0.047</b> | <b>-0.215</b> | <b>11.546</b>                   |

\* The results in bold show the selected model.

**Table 9.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of highest monthly EP values in scenarios without  $\tau$  during training and testing periods at Urmia station.

| Input Scenario                       | Output     | Model | Training     |              |                                 | Testing       |               |                                 |
|--------------------------------------|------------|-------|--------------|--------------|---------------------------------|---------------|---------------|---------------------------------|
|                                      |            |       | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) | CC            | KGE           | RMSE (mm. Month <sup>-1</sup> ) |
| * (1) $EP_{t-1}, EP_{t-2}, EP_{t-3}$ | $EP_t$     | SVM-1 | <b>0.939</b> | <b>0.716</b> | <b>0.369</b>                    | <b>0.188</b>  | <b>0.030</b>  | <b>0.876</b>                    |
| (3) $EP_{t-1}, EP_{t-2}$             | $EP_t$     | SVM-3 | 0.928        | 0.790        | 0.337                           | -0.105        | -0.131        | 1.128                           |
| (5) $EP_{t-1}$                       | $EP_t$     | SVM-5 | 0.538        | 0.213        | 0.717                           | 0.007         | -0.144        | 0.919                           |
| (1) $EP_{t-1}, EP_{t-2}, EP_{t-3}$   | $EP_{t+1}$ | SVM-1 | 0.957        | 0.855        | 0.278                           | -0.371        | -0.464        | 1.070                           |
| (3) $EP_{t-1}, EP_{t-2}$             | $EP_{t+1}$ | SVM-3 | 0.926        | 0.721        | 0.378                           | -0.181        | -0.283        | 1.042                           |
| * (5) $EP_{t-1}$                     | $EP_{t+1}$ | SVM-5 | <b>0.469</b> | <b>0.301</b> | <b>0.788</b>                    | <b>0.272</b>  | <b>0.088</b>  | <b>0.858</b>                    |
| (1) $EP_{t-1}, EP_{t-2}, EP_{t-3}$   | $EP_{t+2}$ | SVM-1 | 0.894        | 0.697        | 0.459                           | -0.612        | -0.651        | 1.115                           |
| (3) $EP_{t-1}, EP_{t-2}$             | $EP_{t+2}$ | SVM-3 | 0.850        | 0.603        | 0.533                           | -0.366        | -0.440        | 0.981                           |
| * (5) $EP_{t-1}$                     | $EP_{t+2}$ | SVM-5 | <b>0.465</b> | <b>0.356</b> | <b>0.846</b>                    | <b>0.046</b>  | <b>-0.009</b> | <b>0.886</b>                    |
| (1) $EP_{t-1}, EP_{t-2}, EP_{t-3}$   | $EP_{t+3}$ | SVM-1 | 0.927        | 0.879        | 0.357                           | -0.299        | -0.300        | 1.245                           |
| (3) $EP_{t-1}, EP_{t-2}$             | $EP_{t+3}$ | SVM-3 | 0.827        | 0.595        | 0.547                           | -0.576        | -0.581        | 1.254                           |
| * (5) $EP_{t-1}$                     | $EP_{t+3}$ | SVM-5 | <b>0.495</b> | <b>0.296</b> | <b>0.788</b>                    | <b>-0.212</b> | <b>-0.293</b> | <b>0.930</b>                    |

\* The results in bold show the selected model.

**Table 10.** Values of the RMSE (mm. month<sup>-1</sup>), KGE and CC criteria for the developed SVM model of highest monthly EP values in scenarios with  $\tau$  during training and testing periods at Urmia station.

| Input Scenario                           | Output     | Model | Training     |              |                                 | Testing       |               |                                 |
|--|------------|-------|--------------|--------------|---------------------------------|---------------|---------------|---------------------------------|
|  |            |       | CC           | KGE          | RMSE (mm. Month <sup>-1</sup> ) | CC            | KGE           | RMSE (mm. Month <sup>-1</sup> ) |
| (2) $\tau, EP_{t-1}, EP_{t-2}, EP_{t-3}$ | $EP_t$     | SVM-2 | 1.000        | 0.988        | 0.015                           | 0.246         | 0.205         | 0.923                           |
| * (4) $\tau, EP_{t-1}, EP_{t-2}$         | $EP_t$     | SVM-4 | <b>0.982</b> | <b>0.862</b> | <b>0.195</b>                    | <b>0.330</b>  | <b>0.236</b>  | <b>0.843</b>                    |
| (6) $\tau, EP_{t-1}$                     | $EP_t$     | SVM-6 | 0.726        | 0.550        | 0.574                           | 0.148         | 0.099         | 0.986                           |
| (2) $\tau, EP_{t-1}, EP_{t-2}, EP_{t-3}$ | $EP_{t+1}$ | SVM-2 | 0.981        | 0.879        | 0.199                           | -0.477        | -0.605        | 1.038                           |
| (4) $\tau, EP_{t-1}, EP_{t-2}$           | $EP_{t+1}$ | SVM-4 | 0.956        | 0.789        | 0.321                           | -0.486        | -0.589        | 1.098                           |
| * (6) $\tau, EP_{t-1}$                   | $EP_{t+1}$ | SVM-6 | <b>0.690</b> | <b>0.522</b> | <b>0.640</b>                    | <b>0.431</b>  | <b>0.237</b>  | <b>0.775</b>                    |
| (2) $\tau, R_{t-1}, R_{t-2}, R_{t-3}$    | $EP_{t+2}$ | SVM-2 | 0.990        | 0.951        | 0.140                           | 0.047         | -0.050        | 0.864                           |
| (4) $\tau, EP_{t-1}, EP_{t-2}$           | $EP_{t+2}$ | SVM-4 | 0.857        | 0.748        | 0.485                           | -0.053        | -0.217        | 0.895                           |
| * (6) $\tau, EP_{t-1}$                   | $EP_{t+2}$ | SVM-6 | <b>0.755</b> | <b>0.646</b> | <b>0.612</b>                    | <b>0.480</b>  | <b>0.245</b>  | <b>0.787</b>                    |
| (2) $\tau, EP_{t-1}, EP_{t-2}, EP_{t-3}$ | $EP_{t+3}$ | SVM-2 | 1.000        | 0.977        | 0.033                           | -0.230        | -0.231        | 1.175                           |
| (4) $\tau, EP_{t-1}, EP_{t-2}$           | $EP_{t+3}$ | SVM-4 | 0.754        | 0.637        | 0.624                           | -0.457        | -0.538        | 1.076                           |
| * (6) $\tau, EP_{t-1}$                   | $EP_{t+3}$ | SVM-6 | <b>0.636</b> | <b>0.517</b> | <b>0.735</b>                    | <b>-0.200</b> | <b>-0.258</b> | <b>1.049</b>                    |

\* The results in bold show the selected model.

#### 4. Conclusions

The present study was prepared with the aim of increasing the accuracy in predicting monthly rainfall (R) and pan evaporation (EP) as two important hydrological processes, by providing a simple solution to determining new inputs for forecasting scenarios. For this purpose, monthly rainfall (R) data of Ardabil station in the period 1976–2019, and pan evaporation (EP) data of Urmia station in the period 1993–2019 were used. Initially, the prediction of the two parameters R and EP for the current and one–three lead times, by determining the different input modes that were a combination of lag times of one, two and three months, was developed with the help of the developed SVM model. Then, in the next step, in order to increase the accuracy of the predictions, application of the simplest input to the scenarios was put on the agenda, and the month number ( $\tau$ ) was added to all scenarios in predicting both the R and EP parameters. The month number was able to greatly improve the prediction accuracy of both the R and EP parameters under the SVM model by entering the scenarios and was able to overcome the complexities within these two hydrological processes that the scenarios were not initially able to solve with high accuracy. This was proven in all current time steps, and the upcoming one, two and three months. Finally, in order to investigate the characteristic of the month number in the SVM model under special conditions such as considering the highest values of the rainfall and pan evaporation time series, it was proved that by using the month number of the SVM model, again, it could improve the accuracy (on average, 17% improvement for rainfall, and 13% for pan evaporation) in almost all time steps. Due to the different conditions of meteorological stations relative to each other, this approach should be considered in different areas, and this is one of the limitations of the present study. Due to the wide range of effects of the two variables studied in the hydrological discussion, the results of the present study can be useful in agricultural sciences and in water management in general and will help owners.

**Author Contributions:** Conceptualization, M.A.J., I.M.-K. and M.A.G.; methodology, M.A.G.; software, S.S.S. and M.A.J.; validation, M.A.J. and I.M.-K.; formal analysis, M.A.G. and V.D.; investigation, S.S.S.; resources, S.S.S. and V.D.; writing—original draft preparation, M.A.J. and M.A.G.; writing—review and editing, I.M.-K. and V.D.; visualization, S.S.S.; supervision, M.A.G. and I.M.-K. 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 of this study are available from the authors upon request.

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

#### References

1. McCuen, R. *Vol. 3 of Hydrologic Analysis and Design*; Prentice Hall: Upper Saddle River, NJ, USA, 2016.
2. Daliakopoulos, I.N.; Coulibaly, P.; Tsanis, I.K. Groundwater level forecasting using artificial neural networks. *J. Hydrol.* **2005**, *309*, 229–240. [[CrossRef](#)]
3. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **2000**, *5*, 115–123. [[CrossRef](#)]
4. Javadi, A.A.; Al-Najjar, M.M. Finite element modeling of contaminant transport in soils including the effect of chemical reactions. *J. Hazard. Mater.* **2007**, *143*, 690–701. [[CrossRef](#)] [[PubMed](#)]
5. Coppola, E.A., Jr.; Rana, A.J.; Poulton, M.M.; Szidarovszky, F.; Uhl, V.W. A neural network model for predicting aquifer water level elevations. *Groundwater* **2005**, *43*, 231–241. [[CrossRef](#)] [[PubMed](#)]
6. Krishna, B.; Satyajai Rao, Y.R.; Vijaya, T. Modelling groundwater levels in an urban coastal aquifer using artificial neural networks. *Hydrol. Process. Int. J.* **2008**, *22*, 1180–1188. [[CrossRef](#)]
7. Basheer, I.A.; Hajmeer, M. Artificial neural networks: Fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [[CrossRef](#)]
8. Chen, S.H.; Jakeman, A.J.; Norton, J.P. Artificial intelligence techniques: An introduction to their use for modelling environmental systems. *Math. Comput. Simul.* **2008**, *78*, 379–400. [[CrossRef](#)]

9. Choubin, B.; Zehtabian, G.; Azareh, A.; Rafiei-Sardooi, E.; Sajedi-Hosseini, F.; Kişi, Ö. Precipitation forecasting using classification and regression trees (CART) model: A comparative study of different approaches. *Environ. Earth Sci.* **2018**, *77*, 1–13. [[CrossRef](#)]
10. Diop, L.; Samadianfard, S.; Bodian, A.; Yaseen, Z.M.; Ghorbani, M.A.; Salimi, H. Annual rainfall forecasting using hybrid artificial intelligence model: Integration of multilayer perceptron with whale optimization algorithm. *Water Resour. Manag.* **2020**, *34*, 733–746. [[CrossRef](#)]
11. Ghorbani, M.A.; Deo, R.C.; Yaseen, Z.M.; Kashani, M.H.; Mohammadi, B. Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: Case study in North Iran. *Theor. Appl. Climatol.* **2018**, *133*, 1119–1131. [[CrossRef](#)]
12. Hadi, S.J.; Abba, S.I.; Sammen, S.S.; Salih, S.Q.; Al-Ansari, N.; Yaseen, Z.M. Non-linear input variable selection approach integrated with non-tuned data intelligence model for streamflow pattern simulation. *IEEE Access* **2019**, *7*, 141533–141548. [[CrossRef](#)]
13. Sammen, S.S.; Mohamed, T.A.; Ghazali, A.H.; El-Shafie, A.H.; Sidek, L.M. Generalized regression neural network for prediction of peak outflow from dam breach. *Water Resour. Manag.* **2017**, *31*, 549–562. [[CrossRef](#)]
14. Tikhamarine, Y.; Souag-Gamane, D.; Ahmed, A.N.; Sammen, S.S.; Kisi, O.; Huang, Y.F.; El-Shafie, A. Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization. *J. Hydrol.* **2020**, *589*, 125133. [[CrossRef](#)]
15. Suparta, W.; Samah, A.A. Rainfall prediction by using ANFIS times series technique in South Tangerang, Indonesia. *Geod. Geodyn.* **2020**, *11*, 411–417. [[CrossRef](#)]
16. Ridwan, W.M.; Sapitang, M.; Aziz, A.; Kushiar, K.F.; Ahmed, A.N.; El-Shafie, A. Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Eng. J.* **2020**, *12*, 1651–1663. [[CrossRef](#)]
17. Yen, M.H.; Liu, D.W.; Hsin, Y.C.; Lin, C.E.; Chen, C.C. Application of the deep learning for the prediction of rainfall in Southern Taiwan. *Sci. Rep.* **2019**, *9*, 1–9. [[CrossRef](#)]
18. Qasem, S.N.; Samadianfard, S.; Kheshtgar, S.; Jarhan, S.; Kisi, O.; Shamsirband, S.; Chau, K.W. Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 177–187. [[CrossRef](#)]
19. Tezel, G.; Buyukyildiz, M. Monthly evaporation forecasting using artificial neural networks and support vector machines. *Theor. Appl. Climatol.* **2016**, *124*, 69–80. [[CrossRef](#)]
20. Malik, A.; Kumar, A.; Kisi, O. Daily pan evaporation estimation using heuristic methods with gamma test. *J. Irrig. Drain. Eng.* **2018**, *144*, 04018023. [[CrossRef](#)]
21. Shiri, J. Evaluation of a neuro-fuzzy technique in estimating pan evaporation values in low-altitude locations. *Meteorol. Appl.* **2019**, *26*, 204–212. [[CrossRef](#)]
22. Kisi, O.; Heddam, S. Evaporation modelling by heuristic regression approaches using only temperature data. *Hydrol. Sci. J.* **2019**, *64*, 653–672. [[CrossRef](#)]
23. Moraes, R.; Valiati, J.F.; Neto, W.P. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Syst. Appl.* **2013**, *40*, 621–633. [[CrossRef](#)]
24. Shin, K.S.; Lee, T.S.; Kim, H.J. An application of support vector machines in bankruptcy prediction model. *Expert Syst. Appl.* **2005**, *28*, 127–135. [[CrossRef](#)]
25. Yoon, H.; Jun, S.C.; Hyun, Y.; Bae, G.O.; Lee, K.K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [[CrossRef](#)]
26. Chen, Q.; Dai, G.; Liu, H. Volume of fluid model for turbulence numerical simulation of stepped spillway overflow. *J. Hydraul. Eng.* **2002**, *128*, 683–688. [[CrossRef](#)]
27. Ehteram, M.; Salih, S.Q.; Yaseen, Z.M. Efficiency evaluation of reverse osmosis desalination plant using hybridized multilayer perceptron with particle swarm optimization. *Environ. Sci. Pollut. Res.* **2020**, *27*, 1–14. [[CrossRef](#)] [[PubMed](#)]
28. Han, S.; Qubo, C.; Meng, H. Parameter selection in SVM with RBF kernel function. In Proceedings of the World Automation Congress, Puerto Vallarta, Mexico, 24–28 June 2012; pp. 1–4.
29. Nguyen, K.A.; Chen, W.; Lin, B.S.; Seeboonruang, U. Using Machine Learning-Based Algorithms to Analyze Erosion Rates of a Watershed in Northern Taiwan. *Sustainability* **2020**, *12*, 2022. [[CrossRef](#)]
30. Shiru, M.S.; Shahid, S.; Park, I. Projection of Water Availability and Sustainability in Nigeria Due to Climate Change. *Sustainability* **2021**, *13*, 6284. [[CrossRef](#)]
31. Yaseen, Z.M.; Al-Juboori, A.M.; Beyaztas, U.; Al-Ansari, N.; Chau, K.W.; Qi, C.; Alig, M.; Salih, S.Q.; Shahid, S. Prediction of evaporation in arid and semi-arid regions: A comparative study using different machine learning models. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 70–89. [[CrossRef](#)]
32. Yaseen, Z.M.; Sulaiman, S.O.; Deo, R.C.; Chau, K.W. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *J. Hydrol.* **2019**, *569*, 387–408. [[CrossRef](#)]
33. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [[CrossRef](#)]