Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2020.24954

177 (2020) 338–349
February

Application of artificial intelligence based and multiple regression techniques for monthly precipitation modeling in coastal and inland stations

Jazuli Abdullahi[a], Anoosheh Iravanian[a,*], Vahid Nourani[a,b], Gozen Elkiran[a]

[a]*Faculty of Civil and Environmental Engineering, Near East University, Near East Boulevard, 99138, Nicosia, North Cyprus, via Mersin 10, Turkey, emails: anoosheh.iravanian@neu.edu.tr (A. Iravanian), jazuli.abdullahi@neu.edu.tr (J. Abdullahi), gozen.elkiran@neu.edu.tr (G. Elkiran)*
[b]*Center of Excellence in Hydro informatics, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran, emails: vnourani@yahoo.com, nourani@tabrizu.ac.ir (V. Nourani)*

Received 18 July 2019; Accepted 22 September 2019

ABSTRACT

The aim of this study was to apply artificial intelligence (AI) based and linear conventional techniques for monthly precipitation modeling in Famagusta (a coastal) station and Nicosia (an inland) station of northern Cyprus. To do so, adaptive neuro fuzzy inference system (ANFIS, as a hybrid technique), support vector regression (as a new AI technique) and multiple linear regression (MLR, as conventional regression technique) were applied in two scenarios. Scenario 1 involved the use of six meteorological parameters as inputs to develop four models from each technique using different input combinations, while Scenario 2 employed the use of only precipitation data at several time lags up to 12 months for the modeling. The results showed that better prediction could be achieved in inland area due to complex and irregular behavior of precipitation in the coastal region. The results also demonstrated that ANFIS models have better performance than models developed by other applied techniques. Scenario 1 models were more efficient and reliable and averagely increased prediction of Scenario 2 models up to 13% for Famagusta station and 18% for Nicosia station in the validation phase. The general results of the study implied that where other meteorological data are not available, precipitation data at previous time steps could sufficiently model monthly precipitation in the study stations.

Keywords: Adaptive neuro fuzzy inference system; Meteorological parameters; Precipitation modeling; Support vector regression

1. Introduction

Being the most significant element of hydrologic water cycle, and for planning and management of hydraulic structures, precipitation plays a vital role and its accurate prediction is of paramount importance. Moreover, due to the complexity, stochastic and nonlinear nature of precipitation time series, the task of its prediction is considerably difficult.

Generally, prediction models for hydro-climatic parameters (such as precipitation) are categorized into two; black box and physically based models. Physical rules are applied for physically based models to properly model all physical processes that are involved in precipitation procedure. Black box models, on the other hand, utilize observed (historical) data to perform further estimations. Computational intelligence and statistical approaches are the basis upon which such black box models are developed. Although for the analysis of actual physics of a phenomenon, conceptual models are dependable methods, but they have some restrictions including data inadequacy for modeling, time consuming, inaccurate results and complexity. Therefore, when predictions are more of concern than physical realization, black

* Corresponding author.

Presented at the 2nd International Conference on Water Problems in the Mediterranean Countries (WPMC-2019), 6–10 May 2019, Nicosia, North Cyprus

box models application could be more useful [1]. Multiple linear regression (MLR) is a conventional method for modeling linear relationship between one or more independent and dependent variables [2]. Toward modeling processes, these kind of linear models basically lose their merit in many areas that are subjected with nonlinearity, dynamism and high complexity in both temporal and spatial scales. In recent time, methods of artificial intelligence (AI) (such as black box methods) showed great accuracy in modeling the dynamic precipitation nature despite the presence of data irregularity, uncertainty and nonlinearity. Comparative researches have shown that application of AI models lead to more efficient and reliable results than physically based models for precipitation predictions [3].

Among AI methods, adaptive neuro fuzzy inference system (ANFIS) is a robust AI model for precipitation predictions, which has both the learning capability of artificial neural network (ANN) and knowledge representation of fuzzy logic. Dastorani et al. [4] applied ANFIS and ANN to model precipitation in the hyper arid climate of Yazd station, Iran which has highly variable and low annual rainfall. Yaseen et al. [5] applied and compared the results of ANFIS with a hybrid ANFIS and firefly optimization algorithm (ANFIS-FFA) for monthly rainfall forecast with one-month lead-time. Apart from ANN and ANFIS, a recent AI model, which was developed on the concept of support vector machine (SVM), is support vector regression (SVR). It is an alternative to ANN and is one of the most useful predicting methods. Mehr et al. [6], and Kisi and Sanikhani [7] applied soft computing methods including SVR, ANFIS and ANN for the prediction of long-term monthly precipitation without climate data. Nourani et al. [1] employed least square support vector machine, ANFIS and ANN to predict precipitation in seven stations located in northern Cyprus (NC).

Despite the significance of precipitation in the hydrologic water cycle and its importance in the planning and management of hydraulic structures, studies (on precipitation) using meteorological parameters, which have profound effect on precipitation, are missing in literature for this study region. In addition, recent studies have shown that on average, precipitation trend decreases whereas annual temperature increases in the NC [8,9]. These emphasized the need for monthly precipitation modeling study with the use of meteorological parameters. Hence, the aim of this study was to evaluate the performances of ANFIS, SVR and MLR techniques in modeling precipitation in Famagusta and Nicosia stations of NC. This was done in two scenarios: (i) Scenario 1 employed the use of different input combinations of meteorological parameters including maximum temperature ($T_{\max}$), minimum temperature ($T_{\min}$), mean temperature (T_{mean}), relative humidity (R_H), wind speed (U_2) and surface pressure (S_p) for the precipitation prediction. (ii) In Scenario 2, precipitation data at previous time steps up to 12 months lag time (to cover seasonality) were used as inputs to the applied techniques.

2. Materials and methods

2.1. Study location and data

Cyprus is the eastern Mediterranean's third largest island and its climate is typically Mediterranean with mild wet winters and hot dry summers, rainfall occurrence mostly is between November and March. Average temperature in NC falls between 5°C and 15°C in the first month of the year (January) and rises higher in July. Recent research showed that through evapotranspiration, about 80% of rainfall water returns to the atmosphere [10]. Data from Famagusta and Nicosia stations were obtained and used in this study (Fig. 1). The Famagusta climate is classified as temperate and warm. It usually constitutes a yearly average rainfall of about 407 mm/year with around 19.3°C average temperature. Being an inland city, the Mediterranean Sea effect is less in Nicosia compared with the other cities in coastal areas. Hence, Nicosia experiences colder winters and hotter summers than the coastal cities. Also there is large difference between day maximum and night minimum temperatures. July and August are the two hottest months and the daytime temperature difference between other cities along the coastline and Nicosia is about 4°C–7°C. For January and February (the coldest months), the difference in daytime temperature of Nicosia is 2°C–3°C less than on the coast [2]. Table 1 gives descriptions of the data used in the study.

As seen in Table 1, in both Famagusta and Nicosia stations, the amount of precipitation can be as low as 0 mm/month which indicates month of no precipitation, but despite being Mediterranean stations, Famagusta has the highest maximum precipitation (245.73 mm/month) than Nicosia (217.80 mm/month) station. This is because, Famagusta is a coastal station surrounded by Mediterranean Sea and the presence of the Sea increases the evaporation bodies of the station, thus results in higher evaporation and subsequently higher precipitation due to cyclic nature of hydrologic water cycle. Fig. 1 shows the location of the study stations.

For the purpose of this study, 36 years monthly data including $T_{\max}$, $T_{\min}$, T_{mean}, R_H, U_2, S_P and P numbering 432 from January, 1983 to December, 2018 for Famagusta and Nicosia stations of NC were obtained and used for the precipitation modeling. Prior to the commencement of the modeling, the monthly average precipitation data were normalized by the equation as follows [1]:

$$P_n = \frac{P_{(t)} - P_{\min(t)}}{P_{\max(t)} - P_{\min(t)}} \tag{1}$$

where P_n is the normalized $P_{(t)}$ data, which has values between 0 and 1 ($0 \le P_n \le 1$), $P_{(t)}$, $P_{\min(t)}$ and $P_{\max(t)}$ are the observed data, minimum and maximum values.

To assess the efficiency and determine the performance of the proposed models, this study endorsed research conducted by Legates and McCabe Jr. [11], which depicted that Nash–Sutcliffe efficiency or determination coefficient (DC) criterion and root mean square error (RMSE) can effectively evaluate any hydro-climatic model, given by:

$$\text{DC} = 1 - \frac{\sum_{i=1}^{N}\left(P_i - \widehat{P}_i\right)^2}{\sum_{i=1}^{N}\left(P_i - \overline{P}\right)^2} \tag{2}$$

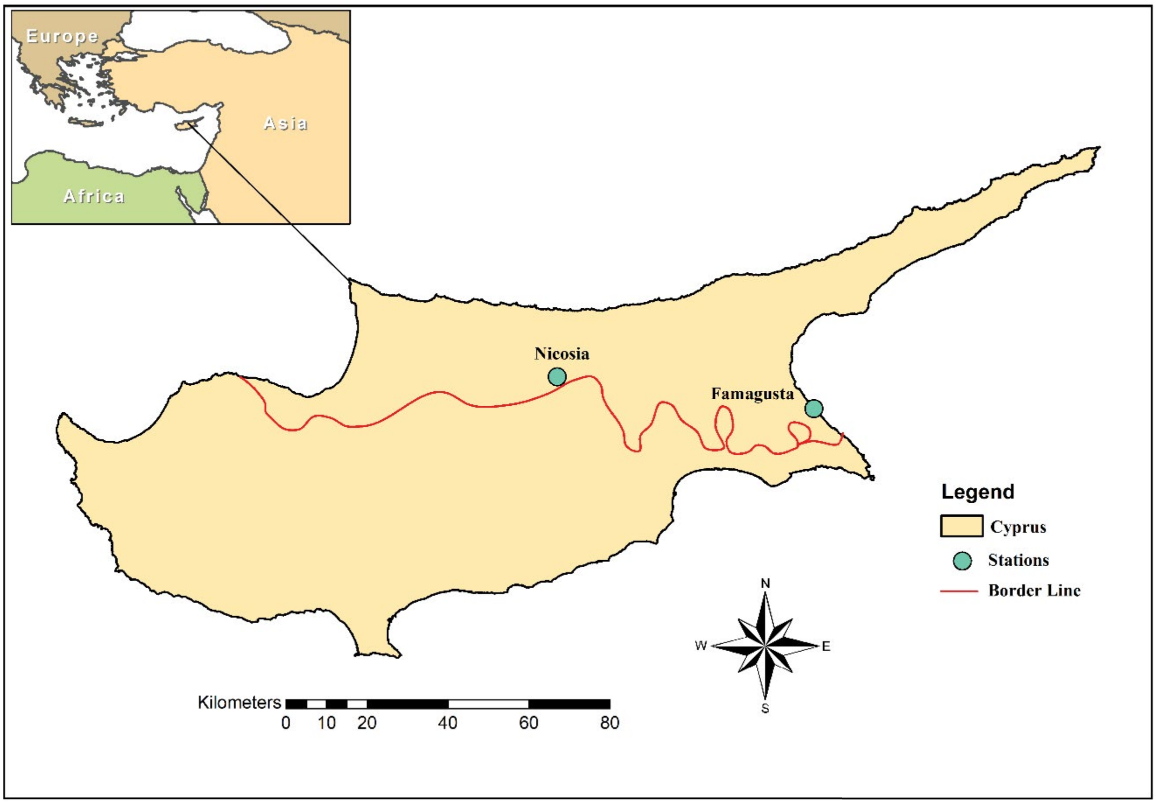

Fig. 1. Location of the study stations.

Table 1
Descriptive statistics of the study data

Station	Parameter	Minimum	Maximum	Mean	Std. deviation
Famagusta	Maximum temperature (T_{max}), °C	12.13	34.11	23.63	6.35
	Minimum temperature (T_{min}), °C	9.03	28.58	19.08	5.46
	Mean temperature (T_{mean}), °C	10.62	31.23	21.28	5.91
	Relative humidity (R_H), %	53.18	75.51	63.90	5.20
	Wind speed (U_2), m/s	1.93	4.95	3.02	0.58
	Surface pressure (S_P), kpa	99.88	101.70	100.75	0.44
	Precipitation (P), mm/month	0	245.73	25.16	31.69
Nicosia	Maximum temperature (T_{max}), °C	11.29	34.76	23.43	6.71
	Minimum temperature (T_{min}), °C	7.41	27.65	17.76	5.60
	Mean temperature (T_{mean}), °C	9.40	30.96	20.42	6.14
	Relative humidity (R_H), %	51.14	77.85	64.51	6.26
	Wind speed (U_2), m/s	2.03	5.03	3.27	0.47
	Surface pressure (S_P), kpa	98.77	100.50	99.58	0.41
	Precipitation (P), mm/month	0	217.80	28.55	34.67

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(P_i - \widehat{P}_i\right)^2}{N}} \tag{3}$$

where N, P_i, $\overline{P}$, and $\hat{P}_i$ are, respectively, the number of observations, observed data, mean of the observed values, and predicted values.

2.2. Validation approach

In this study, a stratified k-fold cross-validation procedure was applied in order to validate the models developed by the used modeling techniques. To implement the k-fold cross-validation, the data samples from Famagusta and Nicosia stations were divided randomly into k = 4 subsamples as demonstrated by Fig. 2. Total sample size

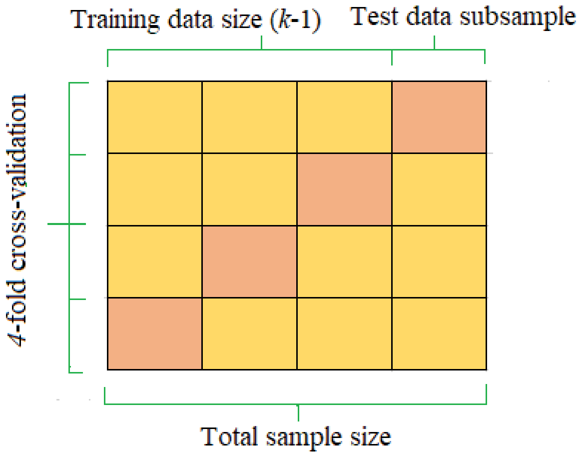

Fig. 2. Schematic illustration of the applied k-fold cross-validation.

was divided by k (4) folds to obtain each random subsample. Single subsample was used to test the model while k–1 (4–1) random subsamples were used for the model training. The process was repeated four times (the number of k-fold subsamples) for different k–1 training subsamples and different single test (validation) subsamples. The advantage of k-fold cross-validation method is that the validation is carried out for each observation exactly once and both the training and validation are carried out using the entire observations [2]. The schematic illustration of the four-fold cross-validation used in this study is shown in Fig. 2.

2.3. Proposed methodology

In this study, ANFIS, SVR and MLR models were developed, trained and validated separately for monthly precipitation modeling using several meteorological parameters as inputs. The proposed approaches were applied via Scenarios 1 and 2.

2.3.1. Scenario 1

This scenario involved the use of meteorological variables as inputs to AI and MLR models for the precipitation modeling. Four models were developed, expressed as:

$$P_t^i = f\left(T_{\max}^i, T_{\min}^i, T_{\text{mean}}^i, R_H^i, U_2^i, S_p^i\right) \quad (4)$$

i implies the station name (Famagusta or Nicosia) and t is observation time, P, $T_{\max}$, $T_{\min}$, T_{mean}, R_H, U_2, S_p were defined in Table 1.

2.3.2. Scenario 2

Instead of using the meteorological variables as inputs, precipitation's own data were used at several time lags (precipitation at previous time steps). Four models were developed from this scenario, expressed as:

$$P_t^i = f\left(P_{(t-1)}^i, P_{(t-2)}^i, P_{(t-3)}^i, P_{(t-4)}^i, P_{(t-5)}^i, P_{(t-6)}^i, P_{(t-12)}^i\right) \quad (5)$$

where $P_{(t-1)}^i$, $P_{(t-2)}^i$, $P_{(t-3)}^i$, $P_{(t-4)}^i$, $P_{(t-5)}^i$ $P_{(t-6)}^i$, $P_{(t-12)}^i$ are the ith station precipitation data at previous time steps t–1, t–2, t–3, t–4, t–5, t–6 and t–12 (or 1, 2, 3, 4, 5, 6 and 12 months ago).

The selection of Scenario 2 for the precipitation modeling was due to the following reasons:

- In modeling precipitation, previous studies such as Nourani et al. [1] and Yaseen et al. [5] have shown that as an auto regression (Markovian) process, precipitation values are more correlated to $P_{(t)}$ at prior time steps ($P_{(t-1)}$, $P_{(t-2)}$ and so on). For this purpose, as inputs for AI models, selection of previous time steps of precipitation values is feasible.
- Due to seasonality of the precipitation phenomenon, $P_{(t-12)}$ was also included as input. This is because of the strong bond (in terms of similarity) that exists between precipitation level at a month of previous year and precipitation values of the same month at current year.

Fig. 3 shows the proposed methodology of the study.

2.4. Adaptive neuro-fuzzy inference system

ANFIS is a robust hybrid system created by the combination of fuzzy system and ANN, which is capable of solving complex nature of relationship [12]. ANFIS is a multi-layer feed-forward neural network that has the ability of integrating fuzzy logic algorithms and the knowledge of ANN, which maps the set of inputs with the output [13]. ANFIS as AI-based model employs hybrid training algorithms which consist of a combination of back propagation and least squares method [14].

The developed ANFIS consists of two inputs of $P(t$–$1)$, $P(t$–$12)$ and one output of $P(t)$ as shown in Fig. 4. The current research employed Takagi –Sugeno-Kang (TSK) fuzzy inference engine for fuzzy operation among different fuzzy inference systems which can be used. The operation of ANFIS to create target function with 2 input vectors of $P(t$–$1)$, $P(t$–$12)$ and first order TSK applied to two fuzzy rules expressed as [15]:

Rule (1): if $\mu(P(t-1))$ is $A1$ and $\mu(P(t-12))$ is $B1$ then $f1 = p1(P(t-1)) + t1(P(t-12)) + r1$

Rule (2): if $\mu(P(t-1))$ is $A2$ and $\mu(P(t-12))$ is $B2$ then $f1 = p2(P(t-1)) + t2(P(t-12)) + r2$

$A1$, $A2$, and $B1$, $B2$ are membership functions parameters, for inputs $P(t$–$1)$ and $P(t$–$12)$ and $p1$, $t1$, $r1$ and $p2$, $t2$, $r2$ are outlet functions' parameters, the formulation and structure of ANFIS followed a five-layer neural network arrangement. The general ANFIS structure is shown in Fig. 4.

2.5. Support vector regression

The concept of SVM learning was introduced by Cortes and Vapnik [16]. It presents a satisfactory approach to the problems of classification, pattern recognition, regression and prediction. SVM-based methods such as SVR is different from many other black box methods, in such a way that instead of minimizing the error between predicted and observed values, the operational risk is considered as the objective function to be minimized. A linear regression is fitted first on the data in SVR, and then to catch the

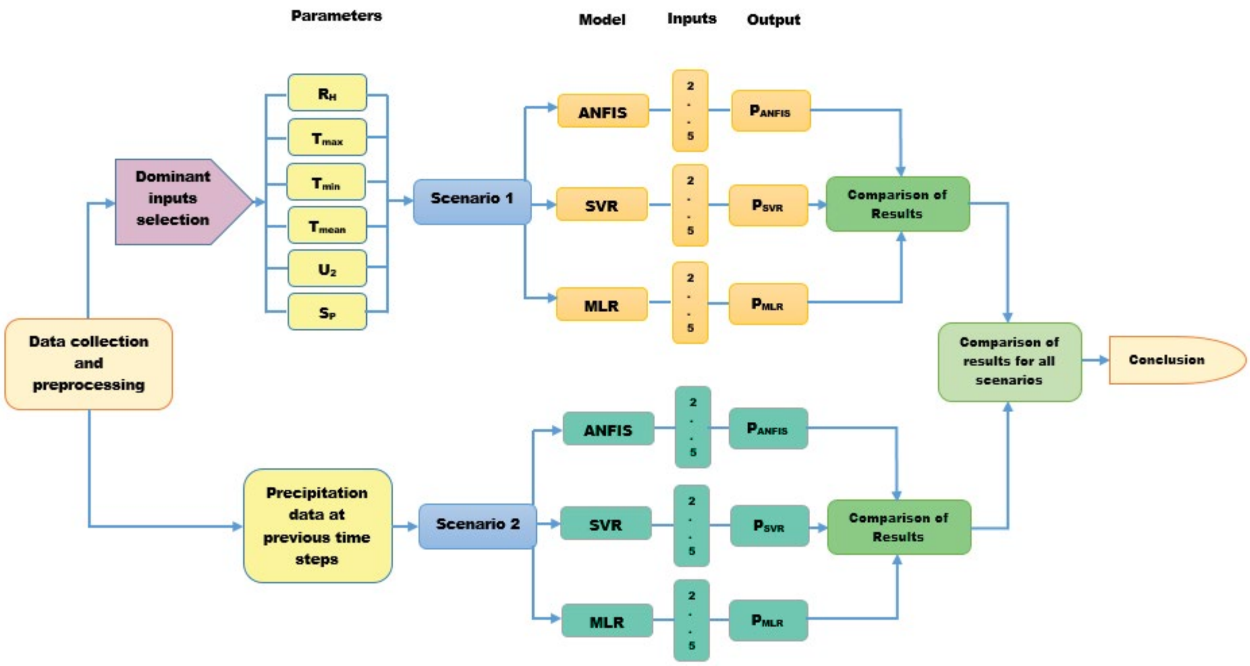

Fig. 3. Proposed methodology of the study.

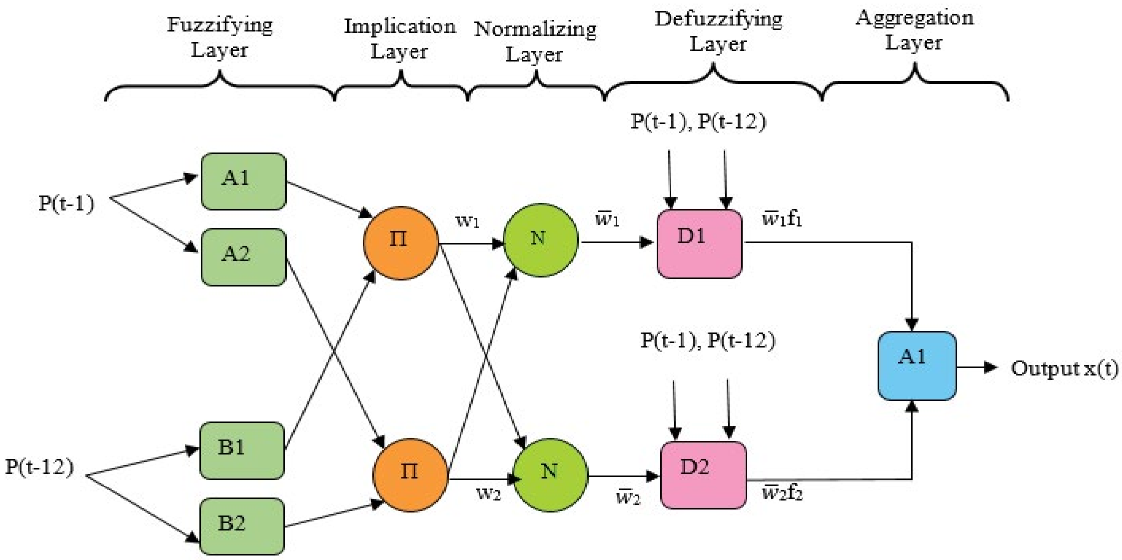

Fig. 4. General ANFIS structure.

nonlinear data pattern, the output goes through a nonlinear kernel.

Given a set of training data, $\{(x_i, d_i\}d_i$ is the actual value, x_i is the input vector and N is the data patterns total number), the general SVR function is [17] as follows:

$$y = f(x) = w\varphi(x_i) + b \tag{6}$$

where $\phi(x_i)$ non-linearly mapped from input vector x which implies feature spaces [18]. Minimization of the objective function and assigning positive values for the slack parameters of ξ and ξ^* may determine regression parameters of b and w [17].

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\left[\sum_i^N \left(\xi_i + \xi_i^*\right)\right] \tag{7}$$

$$\text{Subject to: } \begin{cases} w_i\varphi(x_i) + b_i - d_i \le \varepsilon + \xi_i^* \\ d_i - w_i\varphi(x_i) + b_i \le \varepsilon + \xi_i^* \quad & i = 1,2,\ldots,N \\ \xi_i, \xi_i^* \end{cases} \tag{8}$$

where $\frac{1}{2}\|w\|^2$ is the weights vector norm and the trade-off between the regularized term and the empirical error is determined by C referred to as the regularized constant. ε is equivalent to the approximation accuracy placed within the training data points and is called the tube size. By defining Lagrange multipliers αi and αi^*, optimization problem mentioned can be changed to the dual quadratic optimization problem. After solving the quadratic optimization problem, vector w in Eq. (9) can be computed as [17]:

$$w^* = \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right)\varphi\left(x_i\right) \tag{9}$$

So, SVR can be expressed in the final form as [17]:

$$f\left(x,\alpha_i,\alpha_i^*\right) = \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right)K\left(x,x_i\right) + b \tag{10}$$

b is bias term and non-linear mapping into feature space is performed by $k(x_i, x_j)$ which is the kernel function. Gaussian Radial Basis Function (RBF) kernel is one of the commonly used kernel function as [19]:

$$k\left(x_1,x_2\right) = \exp\left(-\gamma\left\|x_1 - x_2\right\|^2\right) \tag{11}$$

where γ is the kernel parameter. Fig. 5 shows the SVR structure.

2.6. Multiple linear regression

MLR is a famous method of modeling mathematically, the linear relationship between one or more independent variables and dependent variable. In general, the dependent variable y, and n regressor variables may be related by [20]:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_ix_i + \xi \tag{12}$$

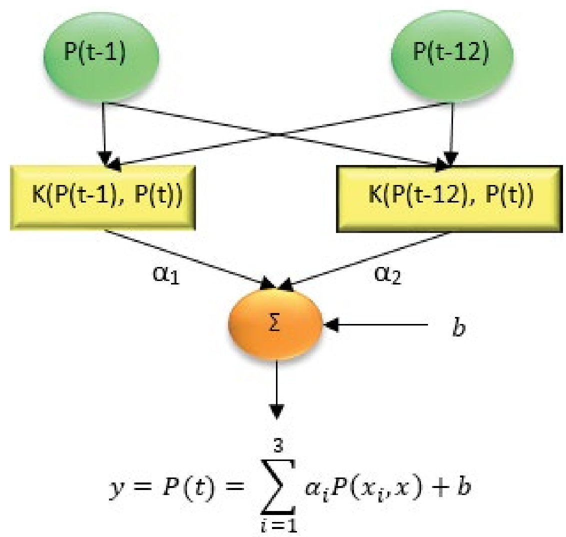

Fig. 5. Structure of SVR.

where x_i is the value of the ith predictor, b_0 is the regression constant, and b_i is the coefficient of the ith predictor and ξ is the error term.

3. Results and discussion

As the study was conducted in Scenarios 1 and 2, the results and discussion section was provided accordingly.

3.1. Results of Scenario 1

Dominant inputs selection is among the difficult and most important aspects of modeling using AI-based techniques. To achieve high performance, the modeling should involve the use of most influential variables as inputs, while less effective and unnecessary variables should be discarded. In some previous studies regarding AI-based modeling, correlation coefficient between potential inputs and target has been employed to select more effective inputs; but as pinpointed by Nourani et al. [21], in a non-linear process, in spite of weak linear correlation, it is possible to have a strong non-linear correlation between the parameters. In view of this, to identify the key input parameters, a single-input single-output ANFIS-based analysis was carried out for the precipitation modeling over the selected stations. The results of dominant inputs selection were ranked from the most influential to least effective parameter and are given in Table 2.

As seen in Table 2 temperatures (T_{max}, T_{min} and T_{mean}) have a major role to play in modeling precipitation, having ranked 1st, 2nd and 6th for Famagusta station and 2nd, 4th and 3rd most influential parameters for Nicosia station. This is because Mediterranean climate is characterized by high temperature (Table 1) which results in higher evapotranspiration that subsequently lead to precipitation. This is in conformity to the study by Payab and Türker [8] whereby between 1900 and 2014, statistical analysis showed an average decreasing trend of monthly precipitation and an increased temperature, which elaborates the effectiveness and correlation that exist between temperature and precipitation. R_H along side T_{max} were found to be the most dominants for Nicosia stations but R_H was the 5th for Famagusta station. The best five most dominant parameters were used for creation of four models for each technique using different input combinations from minimum (two) to maximum (five) inputs.

Based upon the obtained results by dominant inputs selection, the results of the monthly precipitation estimation by ANFIS, SVR and MLR techniques are presented for Famagusta and Nicosia stations of NC.

Sugeno type fuzzy inference algorithm was applied in this study for ANFIS models, where the membership function parameters were calibrated by a set of given input–output data via hybrid optimization algorithm. To achieve the best ANFIS construction, formulation of the ANFIS structures was done through trial and error procedure. Across the two stations for the monthly precipitation modeling, Gaussian-shaped and Triangular membership functions (MFs) were found to be sufficient, while for optimum performance, the number of MFs was determined by examining the modification of training epoch.

For SVR models, the models were created using RBF kernel for the two stations under study. For RBF kernel, the

Table 2
Results of dominant inputs selection

	Famagusta			Nicosia		
Parameter	Training	Validation	Overall	Training	Validation	Overall
$T_{\max}$	1	1	1	2	1	2
$T_{\min}$	2	2	2	4	4	4
T_{mean}	6	6	6	3	3	3
R_H	4	5	5	1	2	1
U_2	3	3	3	6	6	6
S_P	5	4	4	5	5	5

tuning parameters are fewer than two sigmoid and polynomial kernels. In addition, the RBF kernel shows better performance in modeling using SVR technique considering smoothness assumptions [2]. For reliability and efficient precipitation modeling in each station, the RBF kernel's parameters in SVR were tuned. Finally, MLR technique that demonstrates linearly the relationship between independent and dependent parameters was used in this study as well.

Table 3 presents the results of the ANFIS, SVR and MLR models for Scenario 1. It is worth explaining that the results are given for the best output models only. MF-x for ANFIS structure, stand for type of membership function used and number of membership functions. For SVR, RBF is the tuning parameter used in the SVR construction. The MLR structure x-y represents inputs and output parameters used, respectively.

As seen in Table 3, different techniques produced different kind of results at different modeling steps. It can also be seen that in both Famagusta and Nicosia stations, all the applied techniques could produce reliable prediction of precipitation, but models developed by AI techniques surpassed those developed by MLR technique in term of higher DC and lower RMSE. This could be owing to the ability of the AI techniques in dealing with nonlinear and uncertain behavior of precipitation, whereas MLR cannot cope with nonlinear aspect of precipitation. However, among the applied modeling techniques, ANFIS has the best performance which could be due to advantage it has of utilizing both learning algorithm of fuzzy inference system and generalization capability of neural network in a unique framework.

Table 3 also depicts the results of four models developed by AI and MLR techniques with different input combinations. The results show that in both stations, the performances of the developed models increase with increase in number of inputs (except in few occasions such as for Famagusta whereby DCs for M1 and M2 are 0.5532 and 0.5527 by ANFIS technique in the validation phase), but increase in inputs lead to model complication and is time consuming. This could be due to the following reasons:

- Precipitation is uncertain, complex and a nonlinear process that relies on several climatological factors; as a results, models with few meteorological parameters may not capture the desired results as those developed by many parameters.
- Presence of U_2 from M2 to M4 for Famagusta station may have profound effect on the precipitation modeling. As reported by Nourani et al. [22], U_2 might be poor in sensitivity analysis (dominant inputs selection) but its presence with combination of meteorological parameters may increase prediction of hydro-climatological modeling.

Comparison of Table 3 between Famagusta and Nicosia station shows that, with the exception of M1, almost all the applied models provided superior results in Nicosia than in Famagusta station. The better prediction of M1 in coastal station is due to the type of input parameters used. Although, R_H was found to be more correlated with precipitation in the station compared with T_{mean} and $T_{\min}$ according to the given results by dominant inputs analysis (Table 2), but previous studies (including the one by Payab and Türker [9]) show that for the NC, trends of precipitation and temperature are of strong inverse agreement, with the former decreasing while the latter increasing. This clarifies that models developed by using $T_{\max}$ and $T_{\min}$ could be more predictive than the ones based on R_H and $T_{\max}$ inputs. On the other hand, the heat capacity of soil for coastal station is lower than that of inland station. Signifying that the land cools faster and heats faster while in contrast, the ocean heats up and cools down relatively slow. The uncertain cooling and heating behavior of the seacoast makes precipitation process difficult to be predicted by the applied techniques, thus, the models predicted monthly precipitation better in inland station than in coastal station.

The variations of performance of the best model for each technique in the validation phase of Famagusta station are given in Fig. 6 in form of time series and scatter plots.

3.2. Results of Scenario 2

In this subsection of the study, all the modeling procedures and number of models developed are same as in the case of Scenario 1, but precipitation data at previous time steps were used as inputs to the applied techniques. Table 4 shows the results of Scenario 2 precipitation modeling.

As seen in Table 4, similar to results of Scenario 1, models produced by ANFIS have better reliability than those by SVR and MLR techniques. In contrast to Scenario 1, increase in number of inputs does not necessitate increase in models performance. Although it is obvious from Table 4 that M4 (with 5 inputs) provided the best performance across the modeling phases in both stations, M1 (with two inputs) performed better than M2 and M3 (which have three and four inputs). This implies that Markovian process is in strong agreement

Table 3
Results of Scenario 1

Station	Technique	Model	Inputs	Structure	Training		Validation	
					DC	RMSE[a]	DC	RMSE[a]
Famagusta	ANFIS	M1	T_{max}, T_{min}	Gau-2	0.6418	0.0769	0.5532	0.0862
		M2	T_{max}, T_{min}, U_2	Tri-3	0.6398	0.0772	0.5527	0.0862
		M3	T_{max}, T_{min}, U_2, P_S	Tri-4	0.7886	0.05927	0.7193	0.0681
		M4	T_{max}, T_{min}, U_2, P_S, R_H	Gau-5	0.8802	0.0446	0.8143	0.0554
	SVR	M1	T_{max}, T_{min}	RBF	0.5628	0.085	0.5152	0.0898
		M2	T_{max}, T_{min}, U_2	RBF	0.5682	0.0845	0.5245	0.0889
		M3	T_{max}, T_{min}, U_2, P_S	RBF	0.6052	0.0807	0.5801	0.0835
		M4	T_{max}, T_{min}, U_2, P_S, R_H	RBF	0.6047	0.0808	0.581	0.0832
	MLR	M1	T_{max}, T_{min}	2–1	0.5478	0.0863	0.5088	0.0903
		M2	T_{max}, T_{min}, U_2	3–1	0.5521	0.086	0.5213	0.0892
		M3	T_{max}, T_{min}, U_2, P_S	4–1	0.5818	0.0834	0.5788	0.0836
		M4	T_{max}, T_{min}, U_2, P_S, R_H	5–1	0.586	0.0829	0.5796	0.0835
Nicosia	ANFIS	M1	R_H, T_{max}	Tri-2	0.5262	0.1089	0.5194	0.1133
		M2	R_H, T_{max}, T_{mean}	Tri-3	0.7631	0.0801	0.7165	0.0836
		M3	R_H, T_{max}, T_{mean}, T_{min}	Gau-4	0.7669	0.0794	0.7519	0.0782
		M4	R_H, T_{max}, T_{mean}, T_{min}, P_S	Gau-4	0.8675	0.0571	0.8106	0.0716
	SVR	M1	R_H, T_{max}	RBF	0.4984	0.1112	0.4856	0.118
		M2	R_H, T_{max}, T_{mean}	RBF	0.6485	0.0976	0.6324	0.0952
		M3	R_H, T_{max}, T_{mean}, T_{min}	RBF	0.6784	0.0933	0.6516	0.0927
		M4	R_H, T_{max}, T_{mean}, T_{min}, P_S	RBF	0.6815	0.0929	0.6725	0.0899
	MLR	M1	R_H, T_{max}	2–1	0.4766	0.1136	0.4699	0.1198
		M2	R_H, T_{max}, T_{mean}	3–1	0.5712	0.1028	0.5661	0.1084
		M3	R_H, T_{max}, T_{mean}, T_{min}	4–1	0.6034	0.0989	0.6001	0.1041
		M4	R_H, T_{max}, T_{mean}, T_{min}, P_S	5–1	0.6477	0.0932	0.6184	0.1016

RMSE[a]: Since the data were normalized, RMSE has no unit

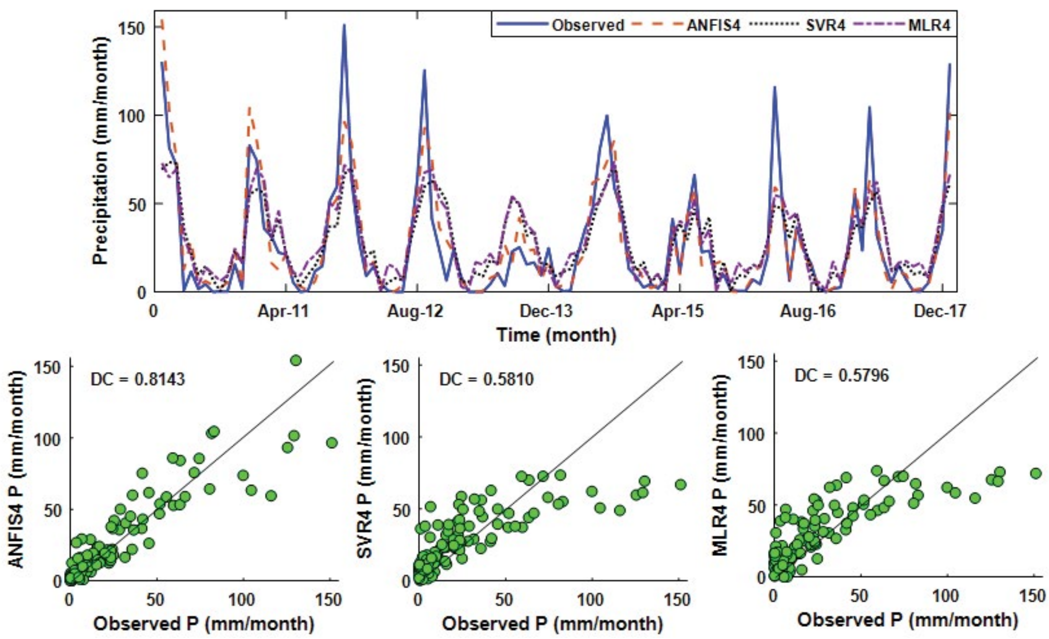

Fig. 6. Best performance models from each technique in the validation phase of Famagusta station in form of (a) time series and (b) scatter plots.

Table 4
Results of Scenario 2

Station	Technique	Model	Inputs	Structure	Training		Validation	
					DC	RMSE[a]	DC	RMSE[a]
Famagusta	ANFIS	M1	$P_{(t-1)}, P_{(t-12)}$	Tri-2	0.5893	0.0824	0.5428	0.0872
		M2	$P_{(t-1)}, P_{(t-3)}, P_{(t-12)}$	Gau-3	0.5478	0.0864	0.5041	0.0908
		M3	$P_{(t-1)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	Tri-4	0.5162	0.0894	0.4666	0.0941
		M4	$P_{(t-1)}, P_{(t-2)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	Gau-5	0.6512	0.0759	0.5982	0.0817
	SVR	M1	$P_{(t-1)}, P_{(t-12)}$	RBF	0.5239	0.0889	0.5087	0.0904
		M2	$P_{(t-1)}, P_{(t-3)}, P_{(t-12)}$	RBF	0.5137	0.0896	0.4856	0.0925
		M3	$P_{(t-1)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	RBF	0.4477	0.0955	0.428	0.0975
		M4	$P_{(t-1)}, P_{(t-2)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	RBF	0.5233	0.08876	0.5109	0.0899
	MLR	M1	$P_{(t-1)}, P_{(t-12)}$	2–1	0.5118	0.0898	0.4886	0.0922
		M2	$P_{(t-1)}, P_{(t-3)}, P_{(t-12)}$	3–1	0.495	0.0914	0.4626	0.0945
		M3	$P_{(t-1)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	4–1	0.4457	0.0957	0.4412	0.0964
		M4	$P_{(t-1)}, P_{(t-2)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	5–1	0.5153	0.0898	0.4899	0.09206
Nicosia	ANFIS	M1	$P_{(t-1)}, P_{(t-12)}$	Gau-2	0.5556	0.1097	0.5548	0.1048
		M2	$P_{(t-1)}, P_{(t-3)}, P_{(t-12)}$	Gau-3	0.5578	0.1044	0.5255	0.1133
		M3	$P_{(t-1)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	Tri-4	0.4987	0.1112	0.483	0.1183
		M4	$P_{(t-1)}, P_{(t-2)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	Tri-4	0.6641	0.0954	0.6259	0.096
	SVR	M1	$P_{(t-1)}, P_{(t-12)}$	RBF	0.4822	0.113	0.4554	0.1214
		M2	$P_{(t-1)}, P_{(t-3)}, P_{(t-12)}$	RBF	0.5404	0.1064	0.4848	0.1181
		M3	$P_{(t-1)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	RBF	0.4584	0.1156	0.4126	0.1261
		M4	$P_{(t-1)}, P_{(t-2)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	RBF	0.5049	0.1105	0.4748	0.1192
	MLR	M1	$P_{(t-1)}, P_{(t-12)}$	2–1	0.4589	0.121	0.4416	0.1173
		M2	$P_{(t-1)}, P_{(t-3)}, P_{(t-12)}$	3–1	0.4886	0.1123	0.443	0.1128
		M3	$P_{(t-1)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	4–1	0.4386	0.1177	0.4063	0.1268
		M4	$P_{(t-1)}, P_{(t-2)}, P_{(t-3)}, P_{(t-6)}, P_{(t-12)}$	5–1	0.4745	0.1138	0.4707	0.1197

RMSE[a]: Since the data were normalized, RMSE has no unit.

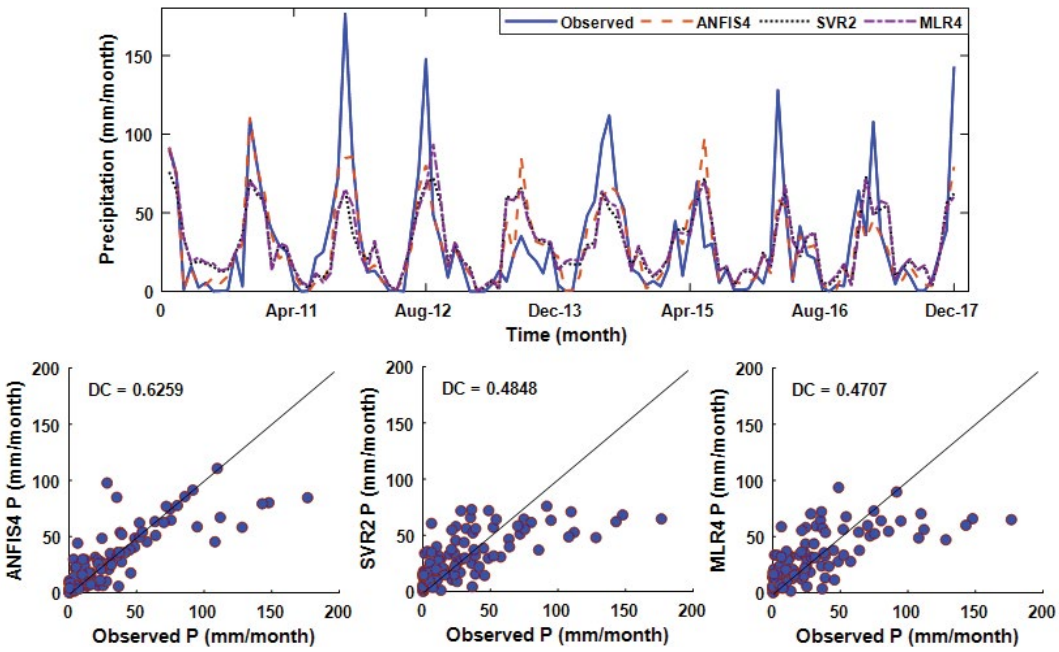

Fig. 7. Best performance models from each technique in the validation phase of Nicosia station in form of (a) time series and (b) scatter plots.

with seasonality of precipitation. For the case of M1 which uses $P_{(t-1)}$ and $P_{(t-12)}$ as inputs, for 12 months lags, there exists a strong bond between precipitation values of the current month and that of the same month in previous year [1]. Moreover, for the case of 1-month lag, higher correlation does exist between current and previous month precipitation. With inclusion of 3 and 6 months lags, the seasonality reduces thereby weakening the Markovian (autoregressive) characteristics of precipitation, which consequently results in lower prediction of precipitation.

As seen in Table 4, the results of the models developed by ANFIS technique in Nicosia station are better than in Famagusta station. This could be because, in contrast to Famagusta station that is located in coastal area, Nicosia station is at central and higher parts of the island. However, the irregular variations of the seacoast increase complexity of precipitation process, thereby making its prediction difficult.

Comparing Scenarios 1 and 2 (Tables 3 and 4), it can be deduced that results of scenario 1 are better and more reliable than results obtained by Scenario 2. This is because independent variables used as inputs for Scenario 1 are important elements in hydrologic water cycle and precipitation being the most significant element could be predicted more efficiently using the factors that influence water cycle. Scenario 1 improved prediction of Scenario 2 models in the validation phase up to 22%, 7% and 9% for Famagusta station, and 18%, 20% and 15% for Nicosia station for ANFIS, SVR and MLR techniques, respectively. However, in spite of less reliability of Scenario 2 with respect to Scenario 1, significant performance was also achieved by Scenario 2 models, thus, where meteorological parameters are not available, precipitation data at previous time steps could be sufficient in modeling precipitation in semiarid Mediterranean climate of Famagusta and Nicosia stations of NC. The variations of performance of the best model vs. observed values for each technique in the validation phase of Nicosia station are given in Fig. 7 in form of time series and scatter plots.

However, to further assess the performance of the individual models for both scenarios, Taylor diagrams were

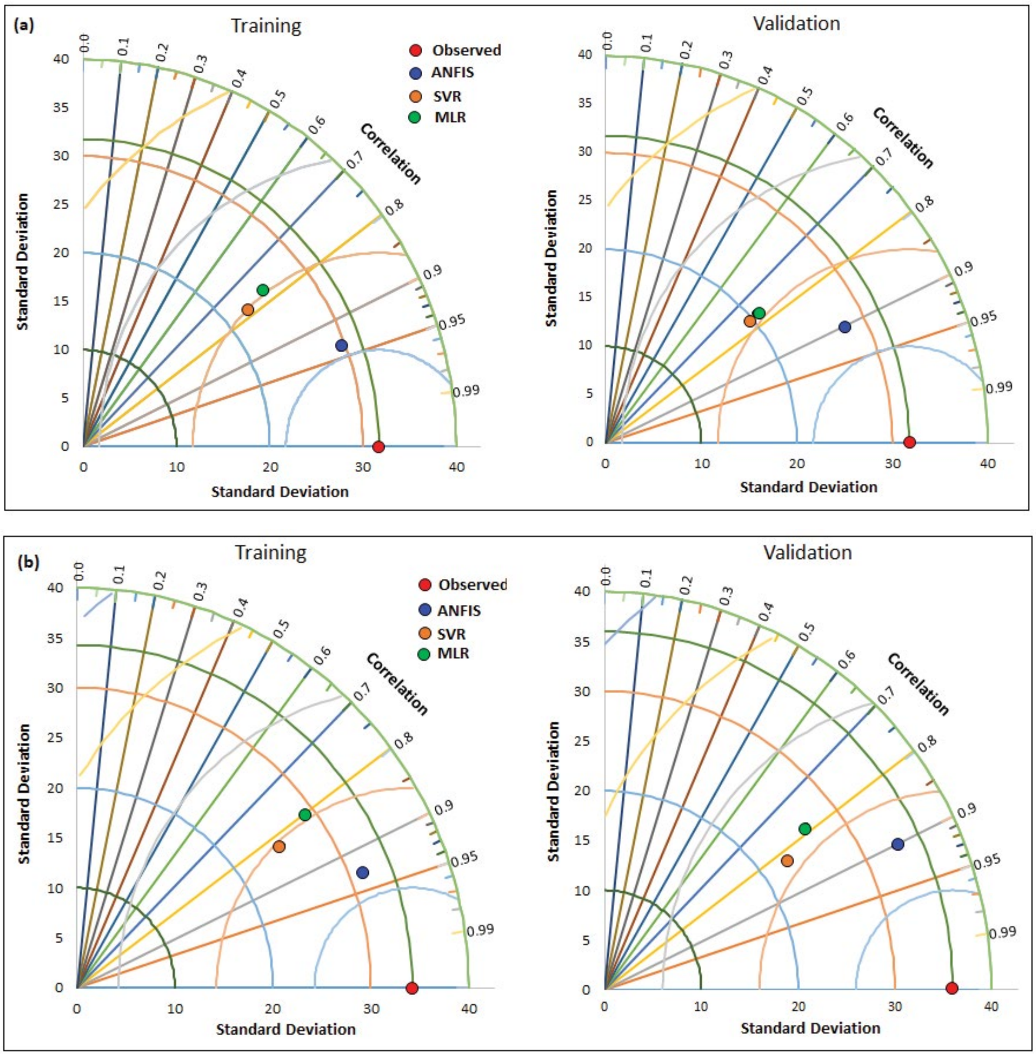

Fig. 8. Taylor diagrams depicting the performance of the applied models in both training and validation phases for the first scenario for (a) Famagusta station and (b) Nicosia station.

plotted. A Taylor diagram summarizes the overall performance of the models by taking into account the variability, pattern correlations, as well as the RMSE between observed data and predictions by the models [6]. In the diagram, the similarity between observed records and predictive models is determined in terms of standard deviation (SD) and correlation coefficient (CC), while RMSE is centered as a measure of distance from observed point (reference point) [23]. In general, if the SD of the observed values surpasses the SD of the predicted values, then underestimation occurs. On the other hand, if the SD of the observed values is lower than the SD of the predicted values, then overestimation occurs [24]. Fig. 8 shows the Taylor diagrams in the training and validation phases of both coastal and inland stations for Scenario 2 for the best models in this study.

Different performances of the models could be seen as demonstrated by Fig. 8. For both Famagusta and Nicosia stations, ANFIS has SD more close to the SD of the observed values and less RMSE which implied better befitting characteristics. Fig. 9 shows Taylor diagrams in the training and validation phases of both coastal and inland stations for Scenario 2 for the best models in this study.

As could be seen in the Taylor diagrams of the second scenario (Fig. 9), there is a wide margin between the SD of the predicted values and SD of the observed values and hence, larger RMSE. This further reaffirmed the earlier presented results that showed higher accuracy of Scenario 1 in comparison with Scenario 2. Apart from SD and RMSE, the accuracy of the models could be ascertained based on the CC values depicted by the Taylor diagrams. For instance, for Famagusta station in the validation phase, the CC value for ANFIS was obtained as 0.9024 for Scenario 1, in comparison with 0.7734 for Scenario 2.

4. Conclusion

In the current study, two AI based and MLR techniques were applied for monthly prediction of precipitation in Famagusta station and Nicosia station of NC. Two scenarios were involved for the modeling purpose. Scenario 1

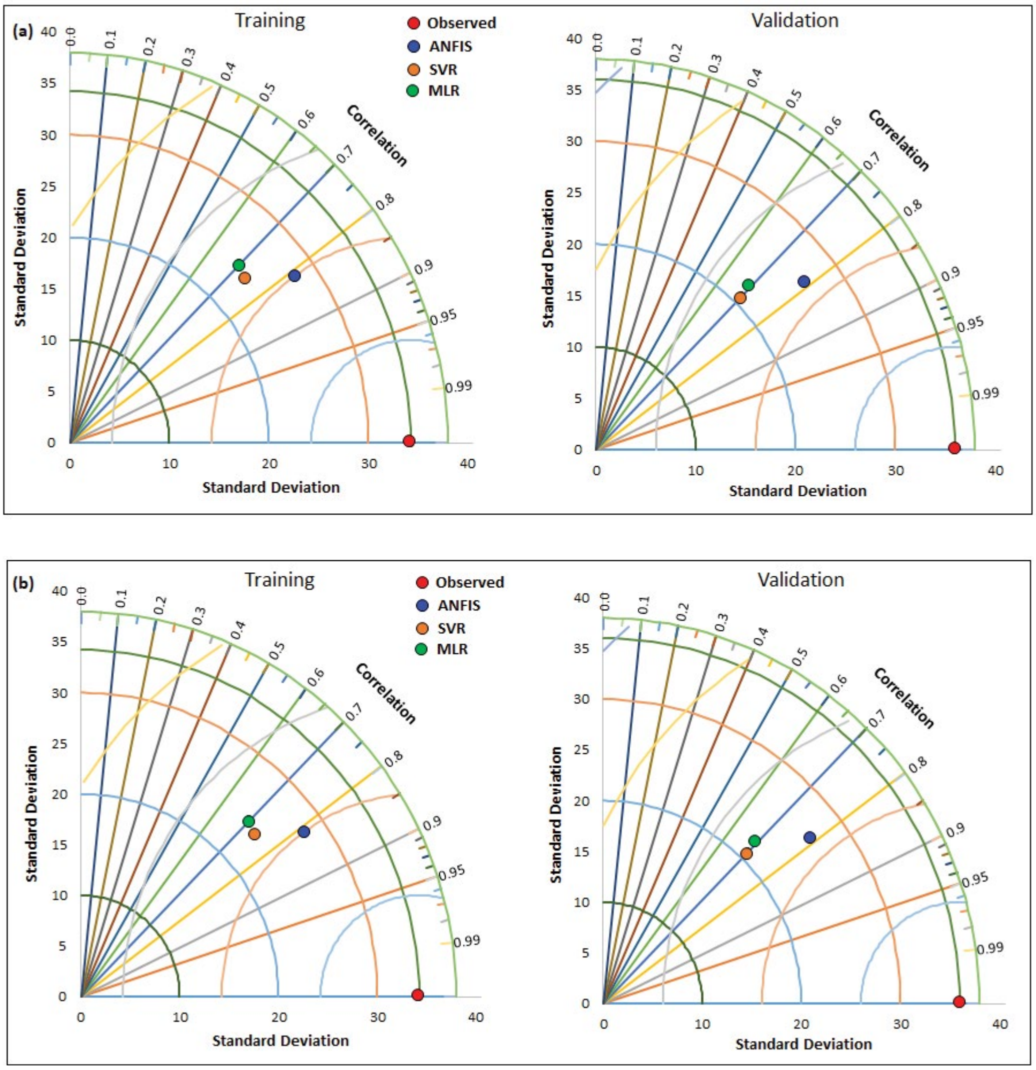

Fig. 9. Taylor diagrams depicting the performance of the applied models in both training and validation phases for the second scenario for (a) Famagusta station and (b) Nicosia station.

evaluated the effectiveness of the applied techniques for the precipitation modeling involving meteorological parameters as inputs using several inputs combination. Before the commencement of the modeling, inputs selection approach was applied to determine dominant parameters for more effective modeling. Scenario 2 involved the use of precipitation data at several time lags using four different inputs combination.

The obtained results showed that maximum temperature is the most dominant among the meteorological parameters due to the location of the study stations to semiarid Mediterranean climate. The results also showed that AI based were superior to MLR-based models, but ANFIS was found to have better performance due to its unique framework of combining neural network and fuzzy inference system generalization capabilities. Scenario 1 models were found to be more efficient than Scenario 2. The overall results of the study demonstrated that the monthly precipitation could be predicted effectively in Famagusta and Nicosia stations of north Cyprus using both the employed scenarios, but for better accuracy and reliability, Scenario 1 is preferable.

With different topography, human activities and vulnerability to drought, varied results could be obtained from other areas and stations within northern Cyprus. Hence, future studies should incorporate other stations that were not considered in this study. In addition, other methods should be tested such as external calibration which uses other station data to validate another, in case of shortage or lack of data in other areas of northern Cyprus.

References

[1] V. Nourani, S. Uzelaltinbulat, F. Sadikoglu, N. Behfar, Artificial intelligence based ensemble modeling for multi-station prediction of precipitation, Atmosphere, 10 (2019) 80.

[2] V. Nourani, G. Elkiran, J. Abdullahi, Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements, J. Hydrol., 577 (2019) 123958.

[3] J. Abbot, J. Marohasy, Application of artificial neural networks to rainfall forecasting in Queensland, Australia, Adv. Atmos. Sci., 29 (2012) 717–730.

[4] M.T. Dastorani, H. Afkhami, H. Sharifidarani, M. Dastorani, Application of ANN and ANFIS models on dryland precipitation prediction (case study: Yazd in Central Iran), J. Appl. Sci., 10 (2010) 2387–2394.

[5] Z.M. Yaseen, M.I. Ghareb, Ebtehaj, H. Bonakdari, R. Siddique, S. Heddam, R. Deo, Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA, Water Resour. Manage., 32 (2018) 105–122.

[6] A.D. Mehr, V. Nourani, V.K. Khosrowshahi, M.A. Ghorbani, A hybrid support vector regression–firefly model for monthly rainfall forecasting, Int. J. Environ. Sci. Technol., 16 (2019) 335–346.

[7] O. Kisi, H. Sanikhani, Prediction of long-term monthly precipitation using several soft computing methods without climatic data, Int. J. Climatol., 35 (2015) 4139–4150.

[8] A.H. Payab, U. Türker, Analyzing temporal–spatial characteristics of drought events in the northern part of Cyprus, Environ. Dev. Sustain., 20 (2018) 1553–1574.

[9] A.H. Payab, U. Türker, Comparison of standardized meteorological indices for drought monitoring at northern part of Cyprus, Environ. Earth Sci., 78 (2019) 309.

[10] G. Elkiran, Z. Ongul, Implications of excessive water withdrawals to the environment of Northern Cyprus, Water Environ. J., 23 (2009) 145–154.

[11] D.R. Legates, G.J. McCabe Jr, Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35 (1999) 233–241.

[12] H.M. Tan, D. Gouwanda, P.E. Poh, Adaptive neural-fuzzy inference system vs. anaerobic digestion model No. 1 for performance prediction of thermophilic anaerobic digestion of palm oil mill effluent, Process Saf. Environ., 117 (2018) 92–99.

[13] A. Solgi, H. Zarei, V. Nourani, R. Bahmani, A new approach to flow simulation using hybrid models, Appl. Water Sci., 7 (2017) 3691–3706.

[14] K.S., Parmar, R. Bhardwaj, River water prediction modeling using neural networks, fuzzy and wavelet coupled model, Water Resour. Manage., 29 (2015) 17–33.

[15] M.A. Sojitra, R.C. Purohit, P.A. Pandya, Comparative study of daily rainfall forecasting models using adaptive-neuro fuzzy inference system (ANFIS), Curr. World Environ., 10 (2015) 529.

[16] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn., 20 (1995) 273–297.

[17] W.C. Wang, D.M. Xu, K.W. Chau, S. Chen, Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD, J. Hydroinform., 15 (2013) 1377–1390.

[18] V. Vapnik, V. Vapnik, Statistical learning theory Wiley, New York, 1998, pp. 156–160.

[19] A.H. Haghiabi, H.M. Azamathulla, A. Parsaie, Prediction of head loss on cascade weir using ANN and SVM, J. Hydrol. Eng., 23 (2017) 102–110.

[20] G. Elkiran, V. Nourani, S.I. Abba, J. Abdullahi, Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river, Global J. Environ. Sci. Manage., 4 (2018) 439–450.

[21] V. Nourani, T.R. Khanghah, A.H. Baghanam, Application of entropy concept for input selection of wavelet-ANN based rainfall-runoff modeling, J. Environ. Inform., 26 (2015) 52–70.

[22] V. Nourani, G. Elkiran, J. Abdullahi, A. Tahsin, Multi-region modeling of daily global solar radiation with artificial intelligence ensemble, Nat. Resour. Res., (2019) 1–22. https://doi.org/10.1007/s11053–018–09450–9.

[23] Z.M. Yaseen, I. Ebtehaj, S. Kim, H. Sanikhani, H. Asadi, M.I. Ghareb, S. Shahid, Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis, Water, 11 (2019) 502.

[24] G. Elkiran, V. Nourani, S.I. Abba, Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach, J. Hydrol., 577 (2019) 123962.