COMPARISON OF DIFFERENT METHODS FOR ESTIMATING MISSING MONTHLY RAINFALL DATA

Abdullah A. Abbas,[1,2] , Nashaat A. Ali[2], Gamal Abozeid[2], and Hassan I. Mohamed[2]

[1] Department of Civil Engineering, Engineering Faculty, Thamar University, Yemen,
E-mail: engaaabbas2011@gmail.com
[2] Department of Civil Engineering, Engineering Faculty, Assiut University, Egypt,
Hassan1@aun.edu.eg

ABSTRACT

Due to its important location and its requirement of hydrological studies for Dhamar watershed, Yemen Republic, this study aims to estimate the missing monthly rainfall data at Dhamar AREA S1 station using different methods. These methods are Normal Ratio method (NR), Inverse Distance Weighting method (IDW), Linear Regression method (LR), and Artificial Neural Networks (ANNs) method. The monthly rainfall data through the period of 1978-1999 were used in different scenarios with 9 different selected surrounding stations. Based on the results of the RMSE, MAE, CC and R^2 error measures, the ANNs method provides the best performance in all periods followed by NR method. The current study found that these methods could be suitable for estimating missing monthly rainfall data in arid and semi-arid regions.

Keywords: Monthly Rainfall, Rain gauge station, Dhamar Governorate, Statistical methods, Missing data, Artificial Neural Network method

1 INTRODUCTION

Water resources planning and management are very important in arid and semi-arid regions such as Yemen. The need of complete and continuous time series data collected over a long period of several meteorological or rain gauge stations for a number of climate parameters such as rainfall are very essential for hydrological studies. However, the problem of incomplete or missing observation data in rainfall is a big challenge. Different methods were described in literature to estimate the daily and or monthly missing rainfall data. Campozano et al (2014) indicated that the infill methods described in literature include three main categories: the deterministic, stochastic, and artificial intelligence methods. Deterministic methods are mathematical models which produce the same output from a given initial condition, while stochastic methods provide probabilistic estimates of the outcome (Campozano et al., 2014). Artificial intelligence methods are modern methods that have a complex mathematical formulation, high cost, and more difficult in implementation. They are effective especially when dealing with non-linear relations (Campozano et al., 2014). The deterministic methods consist of simple arithmetic averaging, inverse distance interpolation, normal ratio method, single best estimator and multiple regressions, etc. The stochastic methods include Kriging and Co-Kriging, etc. The artificial intelligence methods contain Artificial Neural Network (ANNs) and support vector machines (SVM) (Campozano et al., 2014).

Several studies have been conducted using these and other different methods. For instance, Kruizinga and Yperlaan (1978) used different point rainfall estimation methods for the western part of the Netherlands and they concluded that there is no notable difference in performance between the methods used. Tung (1983) conducted a study to estimate missing or ungaged point rainfall in the east side of the Sierra-Nevada mountain using five different methods. The study also developed a method to consider the effect of topographic elevation variation. It concluded that the arithmetic average and inverse-distance methods did not yield desirable results for mountainous regions. Teegavarapu and Chandramouli (2005) Examined several distance-weighted and data-driven methods to estimate missing rainfall data in the state of Kentucky, USA. Their study recommended the best three methods to use in the estimation of missing rainfall data which are coefficient of correlation weighing method, artificial neural network estimation method and kriging estimation methods. De Silva et al (2007) compared four different methods (i.e., aerial precipitation ratio method (APR), the arithmetic mean method (AM), the Normal Ratio (NR) method, and the Inverse Distance Weighting method (IDW) to estimate monthly missing rainfall data in Sri Lankak's zones. The study concluded that IDW, NR, AM and APR are most suitable methods for low country, mid & upcountry intermediate zone, upcountry wet zone, and mid country wet zone respectively. Choge and Regulwar (2013) used Artificial Neural Network Method (ANN) to estimate missing rainfall data in the Maharashtra State, India and concluded that the values predicted by the ANN are reliable to use. They suggested that ANN model can be work for estimation of missing data. Caldera et al (2016) studied seven different methods for filling gaps in daily rainfall data in a mountainous area in Sri Lanka and concluded that it is not possible to name one single method from among the seven methods studied as the most suitable one for all of the stations. It depends on the value of correlation coefficient, either high or low. Sattari et al (2016) evaluated ten different methods for filling in missing monthly rainfall data at six stations located in a southern part of Iran and found that the multiple linear regression method provided a successful estimation of the missing precipitation data.

Thus, in order to estimate missing rainfall data in the rain gauge stations located in Dhamar Governorate, Yemen, this study utilized the most commonly and widely used methods in previous studies. These are: Normal Ratio method (NR), Inverse Distance Weighting method (IDW), Linear Regression method (LR), and Artificial Neural Networks (ANNs) method.

2 STUDY AREA

Dhamar Governorate is located in the middle of western part of Yemen Republic especially in the central highland between latitude of (14^o - 15^o) north and longitude of (43^o 30 - 44^o 50) east (Figure 1). It covers an area of approximately 7587 km^2 in elevation ranges from a few hundred meters to about 3227 m above sea level (Al-Kohlani, 2009; Al Aizari, Lebkiri, Fadli, & Al-Kadasi, 2017; Gibson & Wilkinson, 1995; NIC, 2013). According to the last Census in 2004, the total population for the governorate is about 1,330,108 which expected to be 3,311,033 in 2034 (60% of total increase in 30 years) (CSO, 2004; NIC, 2013).

The governorate is located at an arid and semi-arid climate. According to the classification proposed by UNESCO in 1979, Dhamar is an arid governorate, which is based on the ratio between average annual precipitations (P) and annual reference evaporation (E) (Noaman, 2005). The climate is moderate, although the eastern and central sections of the governorate are likely to be cold during the winter, while the wadis and western slopes are warmer. The average temperatures range from 10 to 19 °C in summer and from 8 to −5 °C in winter. The analysis of collected data shows that the rainfall patterns changed temporally and spatially throughout the governorate. NWRA (2009) implied that the main rainfall periods (March-May) and (July-September) are produced due to the effects of the Red Sea Convergence Zone effect (RSCZ) and inter tropical convergences zone (I.T.C.Z) respectively. It is found that the mean annual rainfall amount varies from < 100 mm in the western coastal areas to about 500 mm in the eastern mountain areas (NWRA, 2009). Due to its important location and the requirement of hydrological studies for Dhamar watershed, the current study is mainly concerned with Dhamar meteorological station in order to estimate its missing rainfall data in the past period. The station is located in the Agricultural Research and Extension Authority (lat.: 1618330 N, long.: 430900 E), and namely Dhamar AREA S1 as shown in Figure 1. The data for the station and other rain gauge stations were collected from different resources to achieve this study.

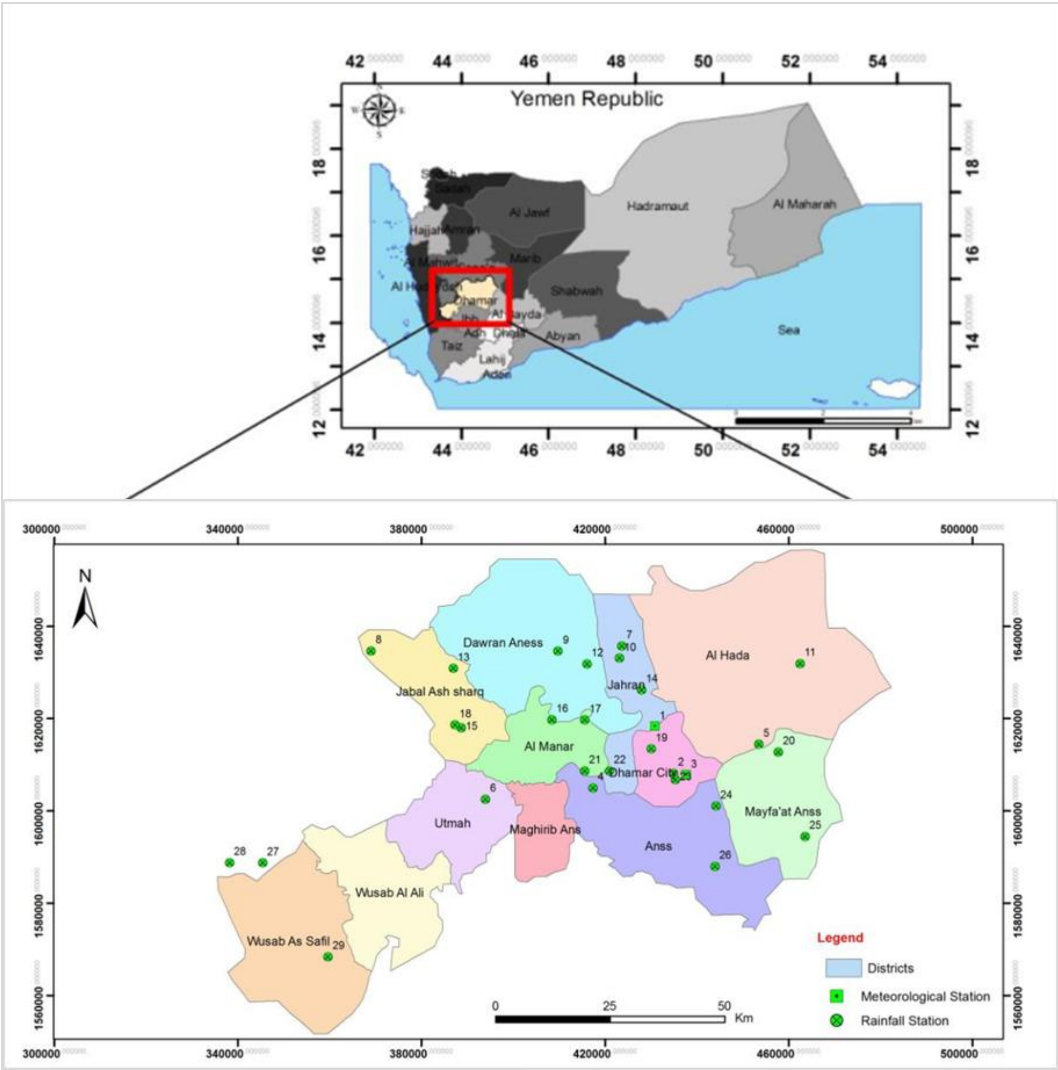

Figure 1. Study area location and the rain gauge station distribution in Dhamar Governorate.

3 DATA USED

The data are obtained from Agricultural Research and Extension Authority (AREA), National Water Resources Authority (NWRA) and Civil Aviation and Meteorology Authority (CAMA) (AREA, 2017; CAMA, 2017; NWRA, 2017). The monthly rainfall data for 29 (S1-S29) meteorological and Rain gauge stations were collected and analyzed. The analyses showed that these stations were distributed in different times with different locations as shown in the map in Figure (1). They have been managed by different Authorities such as, NWRA, AREA, CAMA, and Tehama Development Authority (TDA) (AREA, 2017; CAMA, 2017; NWRA, 2017).

The first observations and measurements of rainfall began in 1970 for different objectives by various consultants who carried out water resources studies in Zabid district for Rima and Zabid wadis from the period 1970 to 1984 (NWRA, 2009). Rain gauge stations have established in the catchment area of these wadis by NWRA. This catchment is located in the eastern mountain areas of wadi Rima. In addition, different rain gauge stations have established on successive years in the governorate. Unfortunately, most of these stations have established in order to achieve some projects and then discontinued after that. Due to that, some rainfall data were not recorded in some of these stations for some years during 1970s, 1980s and 1990s and discontinued after the year 2000 (CES, 2009). The number of stations has increased from about 7 to 17 rainfall and meteorological stations in 1976 to 1983, and then continuously decreased to 11 stations in 1991 to reach 3 stations in 2004.

According to the available and recorded data, the current working stations are Dhamar (AREA) meteorological station (S1), Dhamar City rain gauge station (S2), and CAMA meteorological station (S3). S1 has been established during the beginning of 1970s, while S2 was established and started recording in 2003. There is no available data about the exact date for S3 establishment.

Due to its good location in AREA building, S1 is the best station having continues and reliable records from the early of 1970s till now than other stations, but with missing rainfall data recorded for 1981 Jun-Dec, 1989 Jan-Dec, and 1994-1998 Jan-Dec.

The missing rainfall data for each period will be estimated in different scenarios with different selected stations. To estimate these missing data, the rainfall data through the period of 1978-1999 were used. It is considered as base station, while the other neighbouring rain gauge stations are considering as surrounding stations. Nine surrounding rain gauge stations out of the 29 rain gauge stations are used to realize this objective (the only stations that have full recording). The input data includes two main categories, the rainfall data and the other data (i.e., the distance, the regression coefficient, and the elevation). These data are separated into three main groups based on the period of available rainfall data in surrounding rain gauge stations. These groups range between the periods of 1978-1986, 1980-1989, 1991-1999 as shown in Table 1. The base station is Dhamar (AREA) S1 (Figure 1 and Table 1), while the surrounding rain gauge stations are Wadi-Har (S4), Al-Hajar (S5), Al-Hamal (S13), Dafrd (S14), Ashshirq (S15), Masna'ah (S17), Maram (S20), As-Sanam (S22), and Samah (S24) as shown in Figure 1 and Table 1. Table 1 describes the input data which include the name of base station, period, scenarios, surrounding stations and their numbers. The table also describes the distance, the regression coefficient, and the elevation between the base station and the surrounding stations (Table 1).

Table 1: Input data used to estimate missing rainfall including the base station and Surrounding Stations in different scenarios from 1978 - 1999.

Base Station	Scenario	Surrounding Stations	St. No	Distance (km)	Regression coefficient	Elevation
Dhamar AREA (S1)	1978-1986 (1)	Dhamar (AREA)	S1	0	1	2330
		Dafrd	S14	8.5	0.972	2318
		Ashshirq	S15	42.2	0.534	1500
		Masna'ah	S17	15.3	0.438	2000
		Maram	S20	27.5	0.680	2565
		As-Sanam	S22	13.8	0.486	2400
		Samah	S24	21.8	0.822	2523
	1978-1986 (2)	Dhamar (AREA)	S1	0	1	2330
		Dafrd	S14	8.5	0.972	2318
		Masna'ah	S17	15.3	0.438	2000
		As-Sanam	S22	13.8	0.486	2400
		Samah	S24	21.8	0.822	2523
	1980-1989	Dhamar (AREA)	S1	0	1	2330
		Al-Hajar	S5	22.9	0.75	2640
		Dafrd	S14	8.5	0.79	2318
		Maram	S20	27.5	0.838	2565
	1991-1999	Dhamar (AREA)	S1	0	1	2330
		Wadi-Har	S4	19.0	0.475	2000
		Al-Hamal	S13	45.7	0.507	2000
		Ashshirq	S15	42.2	0.393	1500
		Masna'ah	S17	15.3	0.671	2000

(1) First scenario
(2) Second scenario

Figure 2 shows the input monthly observed rainfall data for the period 1978-1999. Figure 2a shows the monthly observed rainfall data for the first scenario (1978-1986 (1)) at S1 base station and 6 surrounding stations which listed in table (1), S14, S15, S17, S20, S22 and S24. In figure 2b, the monthly observed rainfall data is presented for the second scenario (1978-1986 (2)) at S1 base station and only 4 stations S14, S17, S22 and S24. The S15 and S20 rain gauge stations are excluded in this scenario to study their effects. The monthly observed rainfall data for the period 1980-1989 and 1991-1999 are presented in figures 2c and 2d respectively. It is clear from all these figures that the values of monthly rainfall have the same behavior in all stations during each period. The values of rainfall are nearly close to zero during the period of Oct.-Feb., whereas the heavy amount is occurred during the other months (Mar-Sep.). The maximum rainfall amount is occurred during July. and Aug. months. Therefore, the current studies will focus on estimate the missing rainfall data during the heavy rainfall period using different methods.

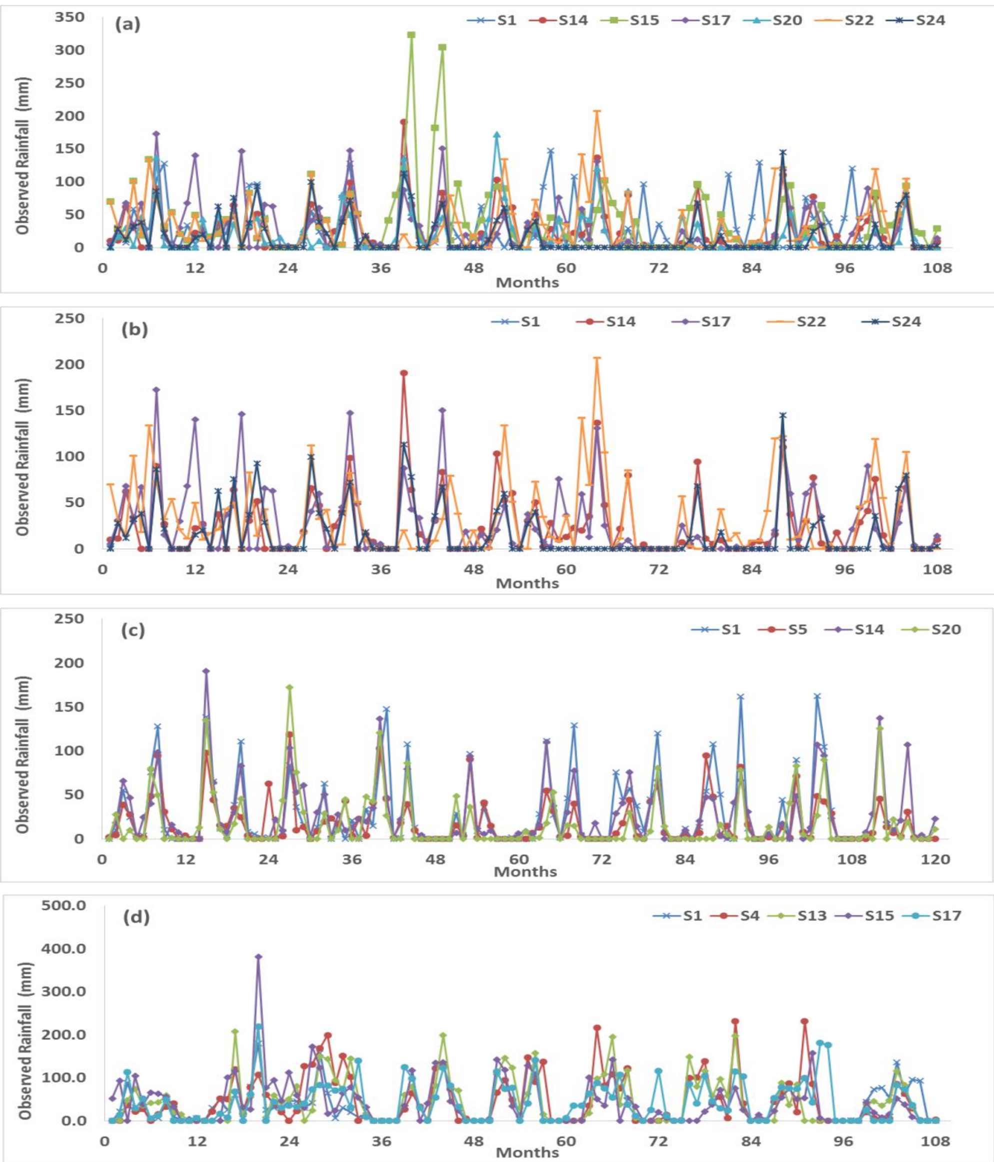

Figure 2: The input monthly observed rainfall data for the period 1978-1999: (a): period 1978-1986 (1); (b): 1978-1986 (2); (c) 1980-1989; (d) 1991-1999 [S4: Wadi-Har, S5: Al-Hajar, S13: Al-Hamal, S14: Dafrd, S15: Ashshirq, S17: Masna'ah, S20: Maram, S22: As-Sanam, S24: Samah.]

4 METHODS

The missing rainfall data for each period is estimated using the following estimation methods (statistical methods and ANNs method).

4.1 Normal Ratio (NR) Method

The normal ratio (NR) method is proposed first in 1952 (Sattari et al., 2016) and then modified year by year in different studies in order to estimate missing rainfall data (Sattari et al., 2016). It is proposed for regions with large station-to-station variation (Tung, 1983). Furthermore, it could be adopted to estimate the missing data if normal annual precipitation of any surrounding rain gauge stations exceeds 10% of the base station (De Silva et al., 2007). This weighs the effect of each neighboring station (Sattari et al., 2016). The NR method have been used in different studies like (Paulhus & Kohler, 1952), (Tung, 1983), (Young, 1992), (De Silva et al., 2007), (Sattari et al., 2016). The estimated missing rainfall observation data at base station x is computed by

$$\mathrm{R}_x = \frac{N_x}{n}\sum_{i=1}^{n}\frac{p_i}{N_i} \qquad (1)$$

In which R_x is the amount of rainfall estimate for the base station x, N_x is the normal annual rainfall of the base station x, p_i is the rainfall values of rain gauges used for estimation, N_i is the normal annual rainfall of the surrounding stations, and n is the total number of surrounding stations used.

4.2 Inverse Distance Weighting (IDW) Method

Inverse distance weighting (IDW) method is most commonly and widely used method for estimation of missing rainfall data (Teegavarapu & Chandramouli, 2005). It was developed and classified as a deterministic method by the U.S. National Weather Service (NWS) in 1972 (Chen & Liu, 2012). Several studies have been used the IDW method such as (Kruizinga & Yperlaan, 1978), (Tung, 1983), (Teegavarapu & Chandramouli, 2005), (De Silva et al., 2007), (Chen & Liu, 2012), (Sattari et al., 2016), (Caldera et al., 2016). In the IDW method, the amount of rainfall to be estimated at a position is a function of rainfall measured and the distance to each neighboring rain gauge stations from the base station (Tung, 1983). The amount of rainfall estimated, R_x at the base station using this method is given by

$$\mathrm{R}_x = \frac{\sum_{i=1}^{n}\frac{1}{d_i^k}p_i}{\sum_{i=1}^{n}\frac{1}{d_i^k}} \qquad (2)$$

Where R_x is the amount of rainfall estimate for the base station x, p_i is the rainfall values of rain gauges used for estimation, d_i is the distance from the location of rain gauge station i to the base station x , and n is the total number of surrounding stations used, and k is referred to as friction distance. Teegavarapu & Chandramouli (2005) and Tung (1983) indicated that k ranges from 1.0 to 6.0 and the mostly commonly used value for k is 2. It have been used as 2 in different studies such as (Tung, 1983), (ZHU & JIA, 2004), (Teegavarapu & Chandramouli, 2005), (De Silva et al., 2007), (Lin & Yu, 2008), and (Caldera et al., 2016). Furthermore, Tung (1983) indicated that the study conducted by Dean and Snyder in 1977 found that for widely spaced gauges in the Piedmont region of the Southeast, the value of *k* of 2 gave the best result. In the current study, the value of k is utilized as 2 to estimate the missing rainfall data using IDW method.

4.3 Linear Regression (LR) Method

Linear regression method can be used to estimate the missing rainfall data in the current study. In this method, the correlation coefficients between the base station and each of the surrounding stations are firstly calculated and then ranked. Then the missing data are estimated using a linear regression equation with the station that has the highest correlation (Caldera et al., 2016). The amount of rainfall estimated, R_x at the base station using this method is given by:

$$\mathrm{R}_x = c_1 p_i \tag{3}$$

Where c_1 is the highest correlation coefficient, R_x is the amount of rainfall estimate for the base station x and p_i is the rainfall values of rain gauges used for estimation.

4.4 Artificial Neural Networks (ANNs) method

Artificial neural networks (ANN) have been applied widely in the past period in the estimation of missing rainfall data and other hydrological studies. Several studies have been conducted to estimate the missing rainfall data and other climate parameters , for example (Kuligowski & Barros, 1998), (Teegavarapu & Chandramouli, 2005), (Coulibaly & Evora, 2007), (Terzi & Çevik, 2012), and (Choge & Regulwar, 2013). ANN have been also trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision, and control systems (Terzi & Çevik, 2012). The definition and procedure of the ANN are adopted from (Beale, Hagan, & Demuth, 1992; Terzi & Çevik, 2012). "Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. A neural network can be trained to perform a particular function by adjusting the values of the connections (weights) between the elements. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. Typically, many such input/target output pairs are used to train a network. Batch training of a network proceeds by making weight and bias changes based on an entire set (batch) of input vectors. Incremental training changes the weights and biases of a network as needed after presentation of each individual input vector".

The backpropagation neural networks, which are also referred to in the literature as feed-forward neural networks or multilayer perceptrons is one of the neural networks which have been commonly used in the estimation of rainfall amount based on surrounding rain gauge stations values. As its described by (Beale et al., 1992), (Kuligowski & Barros, 1998), (Terzi & Çevik, 2012) and other earlier studies, "Feed forward ANNs contain a system of neurons, which are arranged in layers. There is may be one or more hidden layers between the input and output layers. The neurons in each layer are connected to the neurons in a subsequent layer by a weight *w*, which may be adjusted during training. A data pattern including the values *xi* presented at the input layer *i* is propagated forward through the network towards the first hidden layer y_j. Each hidden neuron receives the weighted outputs $w_{ij}x_{ij}$ from the neurons in the previous layer. These are summed to produce a net value, which is then transformed to an output value upon the application of an activation function". A typical three-layer feed-forward ANN is shown in Figure 3. In Figure 3, there are three layers, namely input, hidden and output layers. The input layer neurons are x_1, x_2… x_n which is represented in the current study as rainfall recorded data, the hidden layer neurons are y_1, y_2… y_m, and the output layer neuron is o_1 which is proposed as the missing rainfall value.

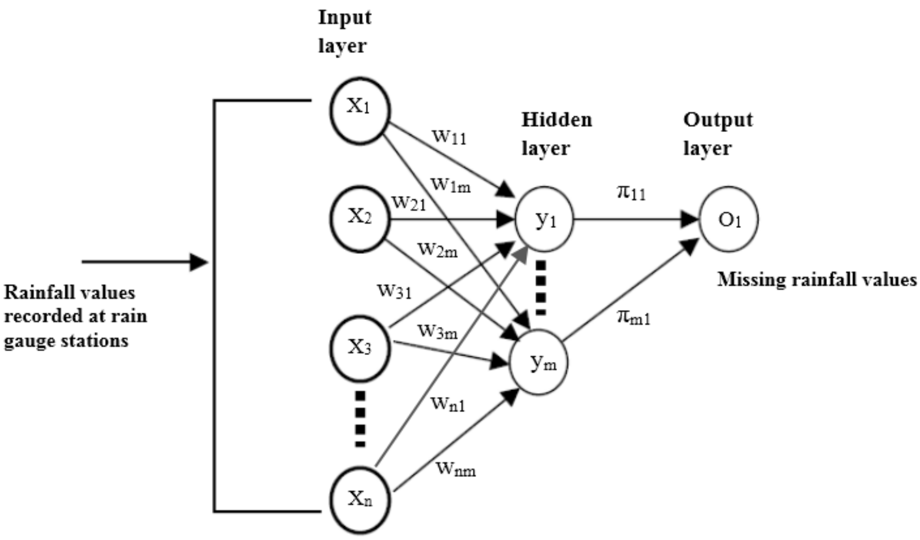

Figure 3. A typical neural network architecture for estimating missing rainfall values at the base station.

5 APPLICATION OF ESTIMATION METHODS

Since the present study aims to estimate the missing monthly rainfall observation data at the base station S1, there are particular steps that applied to realize this objective. Firstly, the 9 surrounding rain gauge stations are used (i.e. S4, S5, S13, S14, S15, S17, S20, S22, and S24) in different periods (i.e. 1978-1986, 1980-1989, and 1991-1999) with two scenarios in the period of 1978-1986 (i.e. 1978-1986 (1), 1978-1986 (2)). Secondly, the data of these stations (i.e. monthly rainfall, distance, regression coefficient) are used as input data to the different estimation methods that mentioned earlier (i.e. NR, IDW, LR, and ANNs). These input data are used to compute the amount of rainfall using equations (1-3). Monthly rainfall data at the base station S1 are supposed to be missing for the purpose of testing and evaluating the estimation methods. For the methods of NR, IDW, and LR all the available data are used, while in ANN method, 70% of the data are used for training the ANNs and the remaining of 30% are used for testing and validating the network. Then, the process and analyses are performed to estimate the missing observation for the base station S1 using each method. In case of ANN, The feed-forward neural network is selected for the analysis. A sample design of ANN could be used in the current study is shown in Figure 3. The input data is proposed to be the monthly rainfall at surrounding stations which are fed into nodes and will pass to the hidden nodes after getting multiplied by weight. The hidden layer neurons are selected using trial and error procedure. The output neurons of the ANN provides the missing value at the stations other than the station of interest. The number of layers used in the present study are three layers, whereas the number of neurons is differ according to the estimation period. It ranges between 6-10 neurons during the training network for the total period of 1978-1999.

Finally, the performance of the estimated value by all methods is compared with the actual value at the base station S1 using commonly used error measures, root mean square error (RMSE), mean absolute error (MAE), correlation coefficient (CC), goodness-of-fit measure criterion, coefficient of determination (R^2). The error measures RMSE and MAE are given by the following equations respectively

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(R_x - R_a)^2} \quad (4)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|R_x - R_a| \quad (5)$$

In which R_x is the estimated or predicted value of rainfall, R_a is the actual value of observation rainfall and m is the total number of observations.

6 RESULTS AND DISCUSSIONS

The results obtained from the methods of NR, IDW, LR, and ANNs are presented in Table 2. Table 2 illustrated the performance of these methods and the comparison using the RMSE, MAE, CC and R^2 error measures. In general, the estimation of missing rainfall by the ANNs method gave the minimum RMSE & MAE, and the maximum CC & R^2 during all the period of 1978-1999 (Table 2). However, in case of statistical methods, there are some methods gave minimum value of errors measures than the other methods during the different periods. For example, during the period of 1978-1986(1), the estimation by IDW method provided minimum RMSE and MAE. However, the CC & R^2 gave maximum values for NR method followed by IDW method at the same period. During the period of 1978-1986(2), the estimation by NR method gave minimum RMSE followed by IDW method. The CC & R^2 gave maximum for the same methods at the same period. The results obtained by MAE errors measure are nearly have the same behavior as shown in Table 2. The same results were obtained for the remaining periods 1980-1989 and 1991-1999 (see Table 2). Overall, based on the results of the RMSE, MAE, CC and R^2 error measures, the ANNs method provides the best performance in all periods followed by NR and IDW methods.

In case of first and second scenario 1978-1986 (1) & 1978-1986 (2) and the difference between them. The results show that the estimation of missing rainfall by the ANNs, NR, IDW and LR methods gave the minimum RMSE & MAE in the second scenario (1978-1986 (2)) compare to the first scenario. This means there is an effect on the values of estimation of missing rainfall due to the use of the 4 rain gauge stations in the second scenario instead of the 6 rain gauge stations. The S15 and S20 rain gauge stations are excluded in this scenario. This effect is probably due to the big difference in distance and elevation of these two rain gauge stations comparing to the other surrounding stations used in the second scenario. Overall, based on the results of the RMSE, MAE error measures, there is no big difference between the first and second scenario. However, the second scenario provides best performance.

Table 2. Performance analyses for different methods of estimating missing monthly rainfall data.

Period	Method	RMSE	MAE	CC	R^2
1978-1986 (1)	NR	31.276	21.072	0.806	0.650
	IDW	29.489	19.257	0.750	0.562
	LR	30.795	21.321	0.743	0.552
	ANNs	14.276	7.308	0.928	0.861
1978-1986 (2)	NR	27.613	20.067	0.793	0.629
	IDW	29.616	19.69	0.746	0.556

	LR	30.795	21.321	0.743	0.552
	ANNs	11.758	4.044	0.969	0.939
1980-1989	NR	26.241	18.249	0.845	0.715
	IDW	29.466	19.93	0.807	0.652
	LR	36.035	23.395	0.735	0.540
	ANNs	7.34	3.238	0.989	0.979
1991-1999	NR	38.101	31.507	0.597	0.357
	IDW	40.891	31.645	0.576	0.332
	LR	43.519	32.413	0.553	0.305
	ANNs	9.197	3.931	0.981	0.963

(1) First scenario
(2) Second scenario

The values of estimated rainfall versus actual rainfall for the different seven methods are presented in Figures 4, 5 and 6. During all periods, there is a well agreement between estimated rainfall and actual rainfall in ANN method followed by NR method.

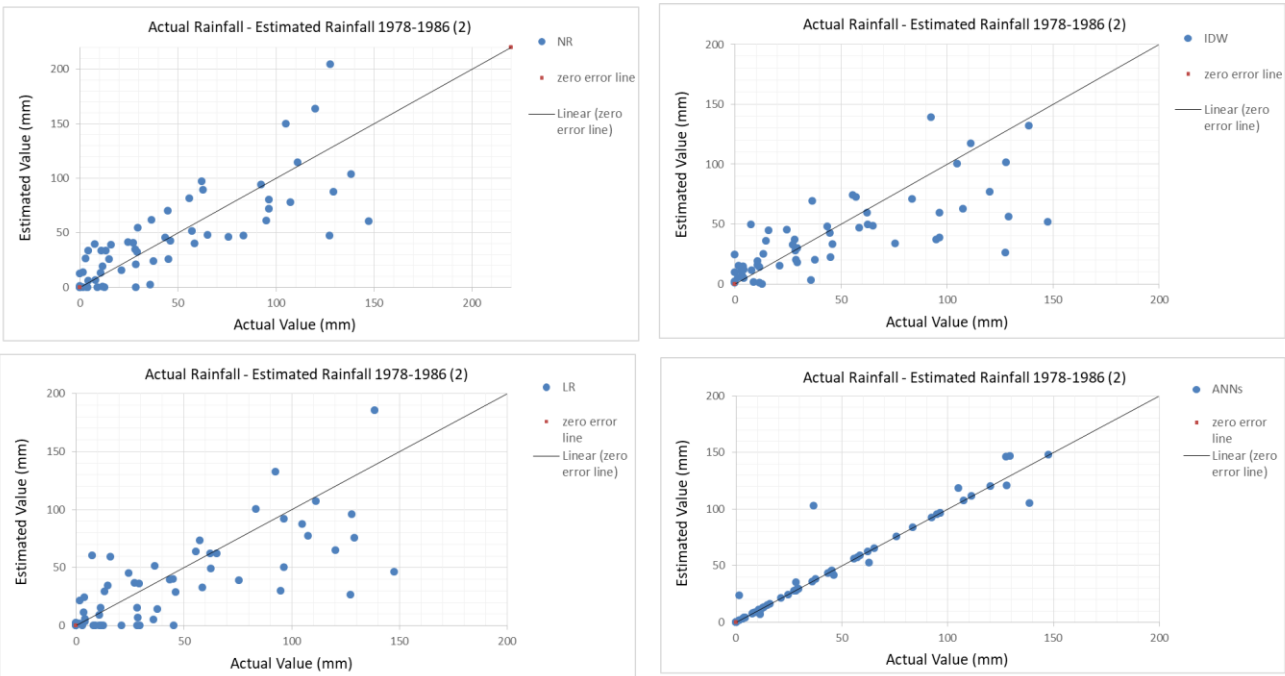

Figure 4. Scatter plot of estimated rainfall versus actual rainfall using the different methods in the period of 1978-1986 (2).

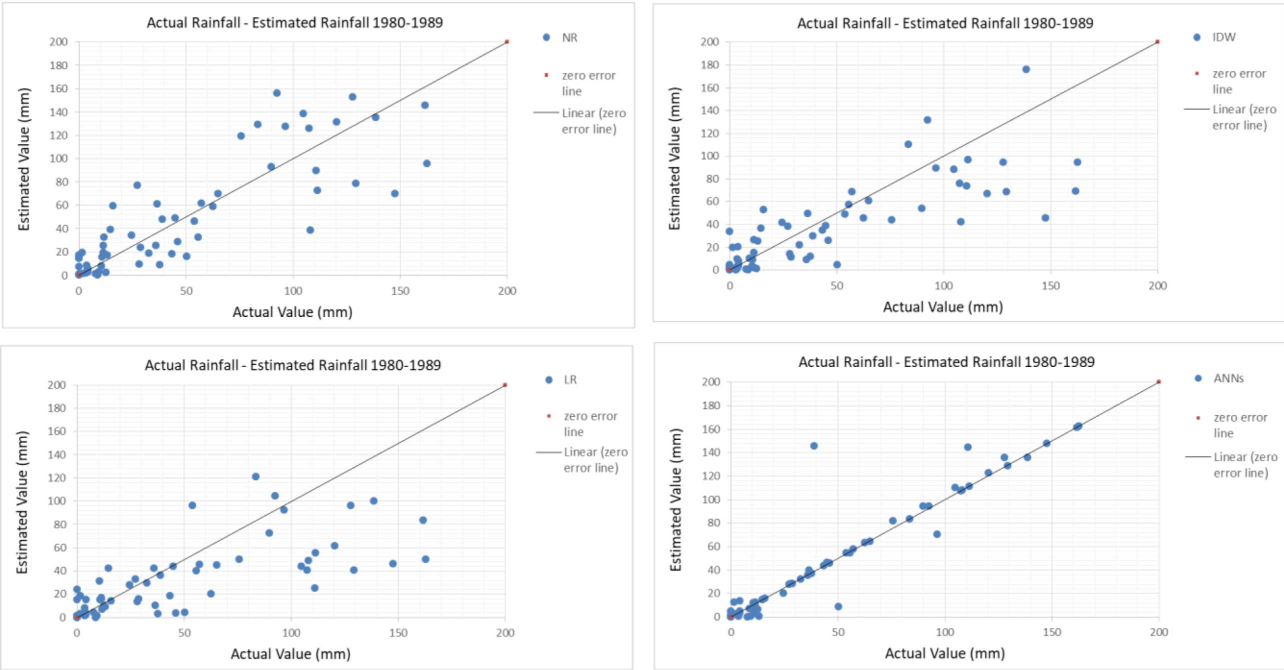

Figure 5. Scatter plot of estimated rainfall versus actual rainfall using the different methods in the period of 1980-1989.

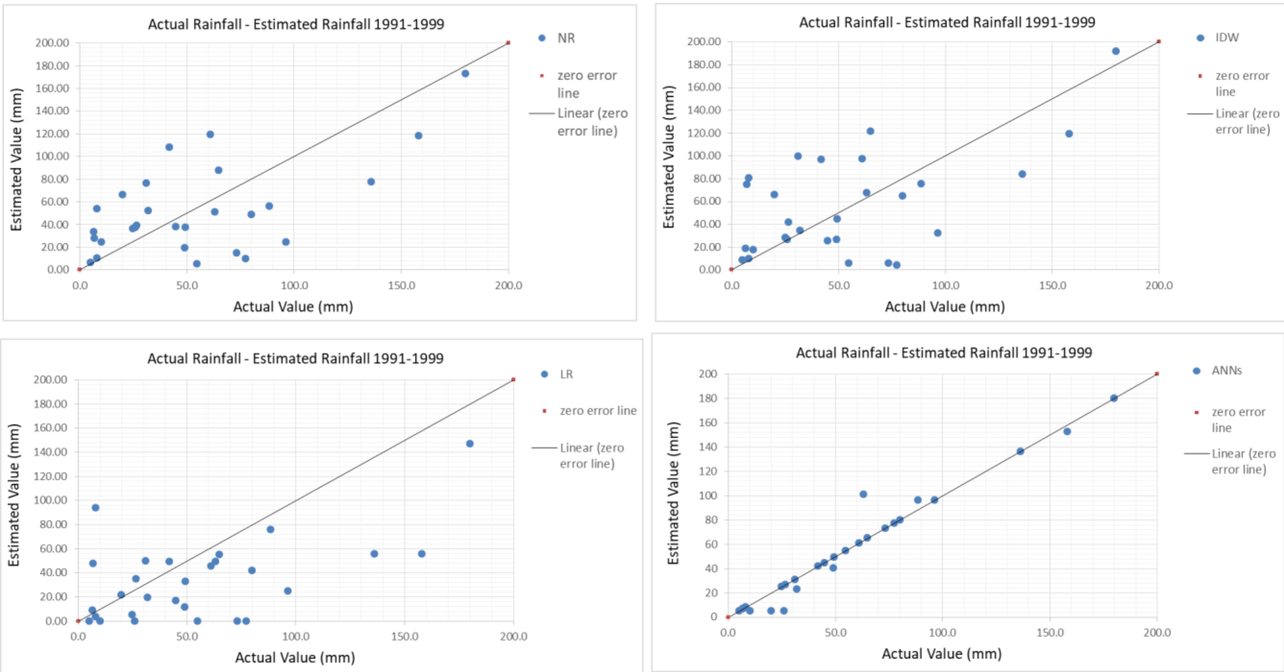

Figure 6. Scatter plot of estimated rainfall versus actual rainfall using the different methods in the period of 1991-1999 .

Moreover, time-series charts and Scatter diagrams produced by the previous methods in the different periods are presented in Figures 7, 8 and 9. These figures include the monthly rainfall data with missing values. The figures illustrate that the ANNs and NR methods provided better results at the S1 station compared with other methods. These methods of estimation are most accurate than other methods of IDW and LR methods which have minimum accuracy among all methods under this study. These results correspond to the results obtained from the performance analysis which presented in Table 2. The figures compare between the actual and estimated values of monthly rainfall at

Dhamar AREA S1 station. It is clear from these figures that the monthly rainfall data are varied and fluctuated from year to year during the different periods.

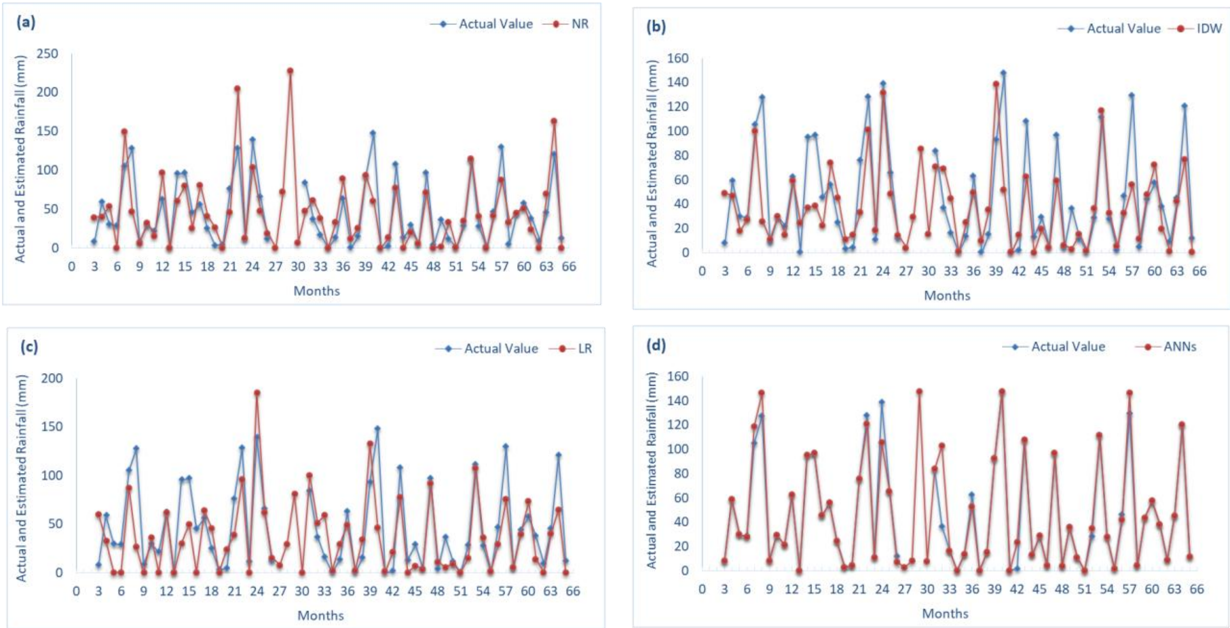

Figure 7. Time series of estimated and actual values of rainfall, during the period 1978-1986 (2) produced by (a) NR, (b) IDW, (c) LR, (d) ANNs.

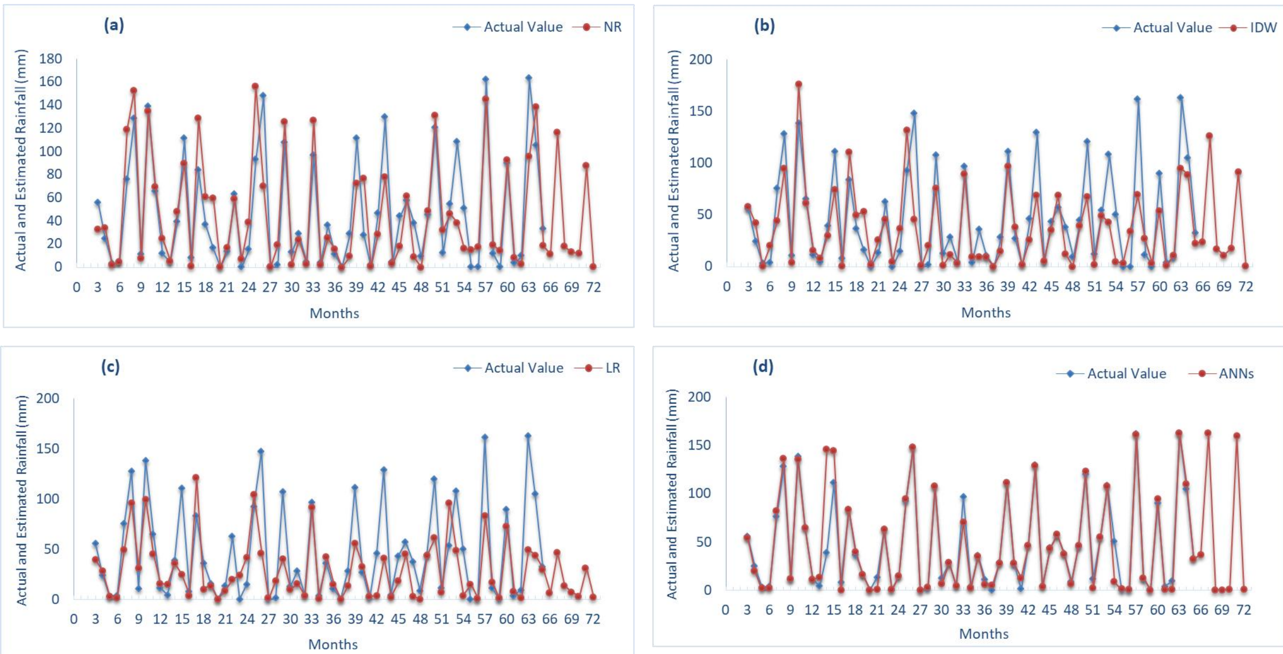

Figure 8. Time series of estimated and actual values of rainfall, during the period 1980-1989 produced by (a) NR, (b) IDW, (c) LR, (d) ANNs.

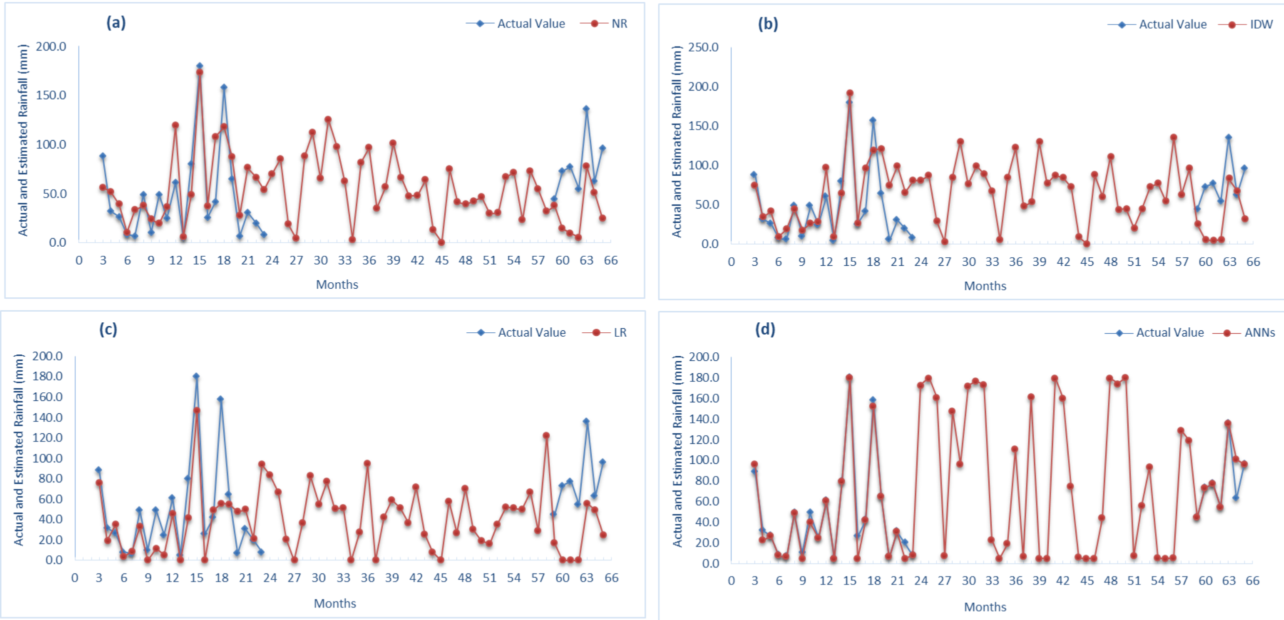

Figure 9. Time series of estimated and actual values of rainfall, during the period 1991-1999 produced by (a) NR, (b) IDW, (c) LR, (d) ANNs.

7 CONCLUSION

This study is conducted to estimate the missing monthly rainfall data at Dhamar AREA (S1) station using 9 surrounding rain gauge stations. Different methods were applied at different periods. These methods were Normal Ratio method (NR), Inverse Distance Weighting method (IDW), Linear Regression method (LR), and Artificial Neural Networks (ANNs) method. The monthly rainfall data through the period of 1978-1999 were used in different scenarios. The computational results verified that the ANNs and NR methods provided better results in estimation of missing rainfall data at the S1 station compared with other methods studied in the current study. The current study found that these methods are suitable for estimating missing monthly rainfall data in arid and semi-arid regions. This study recommends further studies using other statistical methods in order to estimate the missing rainfall data and compare their results with the current methods.

REFERENCES

Adamowski, J., & Karapataki, C. (2010). Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering*, *15*(10), 729-743.

Al-Kohlani, T. A. M. (2009). Geochemistry of thermal waters from Al-Lisi-Isbil geothermal field, Dhamar Governorate, Yemen.

Al Aizari, H., Lebkiri, A., Fadli, M., & Al-Kadasi, F. (2017). Evaluation of Treated Wastewater Quality for Irrigation and its Effect on Soil Infiltration Rate in Dhamar area, Republic of Yemen. *Sun Journal of Chemical and Pharmaceutical Sciences*, *1*(1), 5.

AREA. (2017). climate data and rainfall records. Agricultural Climate Research Unit, Renewable Natural Resources Research Center Agriculture Research and Extension Authority.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (1992). Neural network toolbox™ user's guide. *The Mathworks Inc*.

Caldera, H., Piyathisse, V., & Nandalal, K. (2016). A Comparison of Methods of Estimating Missing Daily Rainfall Data. *Engineer: Journal of the Institution of Engineers, Sri Lanka, 49*(4).

CAMA. (2017). climate data and rainfall records. In c. Data (Ed.). Climate Administration: Civil Aviation and Meteorology Authority, Sanaa.

Campozano, L., Sánchez, E., Avilés, Á., & Samaniego, E. (2014). Evaluation of infilling methods for time series of daily precipitation and temperature: The case of the Ecuadorian Andes. *Maskana*, *5*(1), 99-115.

CES. (2009). Preparation of Water Resource Assessment Plan (WRAP) For Dhamar Plain (RFP No.GSCP/CS/QCBS/2006-05). Dhamar: Consulting Engineering Services (INDIA) PVT. LTD. NEW DELHI SANA'A In Association with AL Amal For Consultants, Yemen.

Chen, F.-W., & Liu, C.-W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment, 10*(3), 209-222.

Choge, H., & Regulwar, D. (2013). Artificial neural network method for estimation of missing data. *International Journal of Advanced Technology in Civil Engineering, 2*, 1-4.

Coulibaly, P., & Evora, N. (2007). Comparison of neural network methods for infilling missing daily weather records. *Journal of Hydrology*, *341*(1-2), 27-41.

CSO. (2004). Statistical Year Book: Population. In C. P. P. o. t. R. o. Y. (2005-2025) (Ed.): Central Statistical Organization, Ministry of Planning and International Cooperation, Sana'a, the Republic of Yemen.

De Silva, R., Dayawansa, N., & Ratnasiri, M. (2007). A comparison of methods used in estimating missing rainfall data. *Journal of agricultural sciences, 3*(2).

Gibson, M., & Wilkinson, T. J. (1995). *The Dhamār Plain, Yemen: A preliminary study of the archaeological landscape.* Paper presented at the Proceedings of the Seminar for Arabian Studies.

Kruizinga, S., & Yperlaan, G. (1978). Spatial interpolation of daily totals of rainfall. *Journal of Hydrology, 36*(1-2), 65-73.

Kuligowski, R., & Barros, A. (1998). *Using Artificial Neural Networks to Estimate Missing Rainfall Data* (Vol. 34).

Lin, X., & Yu, Q. (2008). Study on the spatial interpolation of agroclimatic resources in Chongqing. *J. Anhui Agric*, *36*, 13431-13463.

NIC. (2013). General Profile about Dhamar Governorate. Retrieved 20/12/2017 http://www.yemen-nic.info/, from National Information Center, Sana'a

Noaman, A. (2005). Hydrology of Yemen. http://www.yemenwater.org: Yemenwater.org.

NWRA. (2009). Wadi Zabid and Wadi Rema'a water Quality (Final report). National Water Resources Authority, Sana'a.

NWRA. (2017). Water Resources Database In R. data (Ed.), (pp. Rainfall data records for rainfall stations). http://www.nwrayemen.org/arabic/index.asp National Water Resources Authority, Information Department.

Paulhus, J. L., & Kohler, M. A. (1952). Interpolation of missing precipitation records. *Monthly Weather Review*, *80*(8), 129-133.

Sattari, M.-T., Rezazadeh-Joudi, A., & Kusiak, A. (2016). Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research*, nh2016364.

Teegavarapu, R. S., & Chandramouli, V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, *312*(1-4), 191-206.

Terzi, Ö., & Çevik, E. (2012). RAINFALL ESTIMATION USING ARTIFICIAL NEURAL NETWORK METHOD. *SDU International Journal of Technological Science*, *4*(1).

Tung, Y.-K. (1983). Point rainfall estimation for a mountainous region. *Journal of Hydraulic Engineering*, *109*(10), 1386-1393.

Young, K. C. (1992). A three-way model for interpolating for monthly precipitation values. *Monthly Weather Review*, *120*(11), 2561-2569.

ZHU, H., & JIA, S. (2004). Uncertainty in the spatial interpolation of rainfall data [J]. *Progress in Geography*, *2*.