Nonlinear Rainfall Yearly Prediction based on Autoregressive Artificial Neural Networks Model in Central Jordan using Data Records: 1938-2018

Suhail Sharadqah1, Soraya Mercedes Perez5
Department of Natural Resources and Chemical Engineering
Tafila Technical University, Tafila, Jordan

Ayman M Mansour2
Department of Communication, Electronics and Computer Engineering, Tafila Technical University
Tafila, Jordan

Mohammad A Obeidat3
Department of Electrical Power and Mechatronics Engineering, Tafila Technical University, Tafila, Jordan

Ramiro Marbello4
Faculty of Mines Department of Geosciences and Environment, National university of Colombia-Medellin
Medellin, Colombia

Abstract—Jordan is suffering a chronic water resources shortage. Rainfall is the real input for all water resources in the country. Acceptable accuracy of rainfall prediction is of great importance in order to manage water resources and climate change issues. The actual study include the analysis of time series trends of climate change regards to rainfall parameter. Available rainfall data for five stations from central Jordan where obtained from the Ministry of water and irrigation that cover the interval 1938-2018. Data have been analyzed using Nonlinear Autoregressive Artificial Neural Networks (NAR-ANN) based on Levenberg-Marquardt algorithm. The NAR model tested the rainfall data using one input layer, one hidden layer and one output layer with a different combinations of number of neuron in hidden layer and epochs. The best combination was using 25 neurons and 12 epochs. The classification performance or the quality of result is measured by mean square error (MSE). For all the meteorological stations, the MSE values were negligible ranging between 4.32*10^-4 and 1.83*10^-5. The rainfall prediction result show that forecasting rainfall values in the base of calendar year are almost identical with those estimated for seasonal year when dealing with long record of years. The average predicted rainfall values for the coming ten-year in comparison with long-term rainfall average show; strong decline for Dana station, some decrees for Rashadia station, huge increase in Abur station, and relatively limited change between predicted and long-term average for Busira and Muhai Stations.

Keywords—Jordan; rainfall distribution; time series analyses; Levenberg-Marquardt algorithm; climate change

I. INTRODUCTION

Climate change is now becoming a reality rather than hypothesis. Over the last few decades, the atmospheric concentration of carbon dioxide has increase significantly [1]. This increment beside several other factors induced the average temperature of the planet to increase ≈ 0.2 °C per decade in the past 30 years [2]. The effect of global warming on the variability of rainfall is an important issue. Understanding such variability is essential to a reasonable interpretation of a hydrological cycle response to such increase in earth temperature [3]. The variability of rainfall is a crucial climate component for society, environment, agriculture, and over all water management plan. Increasing precipitation variability can produce a long series of effects, from reducing the productivity of agriculture to affecting the growth of children [4], [5].

Jordan is considered one of the poorest countries in water resources. Therefore, water resources related issues are always present a source of concern and at the same time a source of interest. Among these issues, we could specify the climate change, groundwater over abstraction, water quality deterioration. Groundwater depletion and precipitation decreasing or precipitation time shifting. As a result, climate change studies are increasingly important [6]. There are historical indications of climate change, or at least there are apparently some areas that used to have Ecosystems that need more water sources than what is available now days [7], [8], [9]. One of the good examples that show these supposed climatic changes is the murals in the Umayyad palaces scattered in the Jordanian deserts. The murals present landscape, fauna and flora that require much water than available resources now days [10], [11]. The Dead Sea is another example of the supposed climate change. The Dead Sea suffers from a continuous decline in its level, and consequently a significant decrease in its area, as well as the disappearance of a number of streams and small rivers that used to discharge to it [12], [13], [14]. The official water resources policy pays great emphasis on climate change, because the country's future and prosperity are related to the abundance of water resources and their adequacy to meet current or expected future requirements [6]. It should be noted that in addition to climate change, the large and irregular increase in population, whether due to natural reproduction or migrations from neighboring countries, complicates the water resources management efforts in Jordan. The rapid growth of population is partially attribute of massive immigration from neighboring countries due to wars and insecurities [15], [16].
Climate change and related issues are of great concern to government agencies, scientific entities and even public masses. By this means, until 2014 the Jordanian ministry of Environment had patronage at least three National Communications on Climate Change. Many scientific articles that deal with climate change, its indicators and impacts have been published especially since 2004 [17], [18].

For common population in Jordan, climate change is of increasing concern. The Dead Sea tragedy presents a chock hop for public awareness regards to climate change. Only in approximately one quarter of hour, a heavy rainfall produced huge flash flood that swept more than 20 people to their death on October 25th, 2018. On November 10th of the same year another flashflood killed and injured more than 40 people and inundated more than 1000 people in Maan area [19]. This study represent a sample of the scientific response to the massive social worry regards to climate change. Where a good prediction for the pattern and distribution of the future rainfalls will be very helpful for safe management of water resources and anticipate some rainfall related disasters.

In order to forecast the rainfall many methods have been used. Among these methods; simple regression analysis, autoregressive integrated moving average, and exponential smoothing techniques. The accuracy of these methods still in debate until the moment [20], [21], [22]. The development of computing techniques and capabilities enhance the using of certain methods in climate parameters forecasting. The Artificial neural networks is good example of the new techniques where in many studies present a good accuracy [23], [24].

Recently, Artificial intelligent is used in many applications such as power [25], health [26-31], communication [32], text classification [33], [34], texture classification [35],[36], and optimization [37]. Technology and AI applications can be applied in many different sectors and industries to generate maximum production from the operational front. Artificial neural networks are one of the recent trends used in AI applications such as communication [38], wind power prediction [39],[40], text classification [41], civil engineering [42], health [43], image processing [44], climate prediction [45],[46],[47], and power load forecasting [48]. One of the most important features of ANN is its ability to recognize time series data and predict data with high efficiency compared to other methods, especially nonlinear relationships. Predicting the amount of rain that will fall in a given area for a long term is a difficult problem that is still been studied.

There are several theories used in the literature to make long-term prediction, such as linear and nonlinear techniques. Common nonlinear methods are Neural Networks, Support Vector Machine (SVM) and fuzzy logic. Artificial Neural Networks (ANN) has been successfully applied to solve some complex practical problems. ANN can discover and learn relationships between sets of data and link them together. ANN has great capabilities in dealing with nonlinear prediction problems. In addition to that, it has the ability to handle massive numbers of variables. There are several types of them, as mentioned in the following references. After training, it can be used to predict the rainfall.

There has been an increasing interest in the scientific literature regarding accurate prediction in the case of linear and nonlinear systems using artificial neural networks. There are many practical applications in this field. In this scientific paper, the possibility of obtaining annual forecasts of precipitation quantities through neural networks is studied and analyzed. NAR neural networks based on Levenberg-Marquardt algorithm is used here. Prediction results showed a high degree of accuracy of long-term forecasts.

The aims of the actual study are the followings:

- Investigate the availability of rainfall data for a conscious predation study.
- Predict the future rainfall quantity on the base of seasonal year and calendar year in the study area.
- Check the accuracy of ANN techniques for climate forecasting in the study area.
- Observe the climate change pattern related to rainfall parameter in the study area.

II. SYSTEM DESIGN AND METHODOLOGY

In this section system design for rainfall amount forecasting is presented. In addition, the used methodology is explained.

A. Study Area and Data Collection

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or head unless they are unavoidable. Five rainfall stations were selected in the central region of the Hashemite Kingdom of Jordan (Fig. 1). One of these stations (Mohi) is located in Al-Karak Governorate and the others are in Al-Tafila Governorate. All these stations are located at an altitude of more than one thousand meters above sea level. The areas represented by these stations are the most populated centers in the two governorates.

A large part of the population in these areas practices agriculture, and therefore changes in rainfall levels will have a great impact on the lives of citizens in addition to their impact on the water balance and water resources in general.

![Fig. 1. Study Area Map shows the Spatial Distribution of the Five Rainfall Stationes.](image-url)
Available rainfall data for the five metrological stations were obtained from the Ministry of Water and Irrigation. The data cover time span extend since the station started working till the 2018 rain season. For some stations the first data back to 1937 where for other the first data backs to 1969 (Table I). The density of data is also a varying issue. Some stations have a good record of data while others have some temporal gaps.

The data include the daily rainfall quantities in mm. The collected data from the five different locations are combined to form one central database. Fig. 2 shows the elevation map of the study area with station ID’s.

Table II shows an example of a data from one of the sites (Dana). Fig. 3 shows plot of real rainfall data of the same station from 1945-2010.

B. Data Preprocessing Phase

The pre-process of the data, including data cleaning and missing data treatment. The unnecessary information for rainfall amount model such as Object ID is removed from the database. Then the missing rain in a certain date is replaced by the average value of the rain on the same year. The calendar rainfall values are calculated by summing the daily values from January 1st to December 31st of the calendar year.

| Station Name | Observation period | Comments         |
|--------------|--------------------|------------------|
| Muhi         | 1968-2017          | Sporadic short gaps |
| Tafila       | 1938-2018          | Data available since October 1937 |
| Busira       | 1938-2018          | Data available since October 1937 |
| Rashadia     | 1970-2018          | Sporadic short gaps |
| Dana         | 1946-2011          | big gap between 2005-2009 |

TABLE I. TEMPORAL COVERAGE OF RAINFALL DATA FOR THE FIVE STATIONS

Fig. 2. Elevation Map of the Study Area.

The seasonal rainfall values are calculated by summing the daily values from September 1st of a year to August 31st of the followed year. In some studies they suppose that the rain fall season starts in October 1st to September 30 of the next year. Practically this is not issue for the actual study area, where the rainfall from June to end of September is almost zero [16]. Fig. 4 shows the long-term monthly distribution of rainfall in Tafila governorate where practically no considerable rainfall is recorded during June, July, August and September.

C. Nonlinear Autoregressive Artificial Neural Networks Model

The template is designed so that author affiliations are not The NAR network is a feed forward neural network with three layers; input, hidden and output layers. NAR network is used to solve a time series problem. NAR neural network uses historical data of the rainfall in order to do the prediction. The topology of a NAR network is shown in Fig. 5. The number of delays and the number of neurons in the hidden layer are adjustable. These numbers are optimized through trial-and-error testing in order to get accurate model responses.

Fig. 3. Plot of Real Rainfall Data of Dana Station from 1945-2010.

Fig. 4. Long Term Monthly Distribution of Rainfall in Tafila Governorate.
Future values of time series values $y(n)$ are forecasted based on its delays values as shown in (1).

$$y(n) = \sum_{j=1}^{m} w_j \cdot \left[ f \left( \sum_{i=1}^{k} w_{ij} y(n-i) + w_{0j} \right) \right] + w_0 + \varepsilon \quad (1)$$

Where $k$ is the number of delays, $m$ is the number of neurons in a hidden layers with activation function $f$, and $w_{ij}$ is the weight of the connection between the input $I$ and the hidden neuron $j$, $w_j$ is the weight between the hidden neuron $j$ and the output layer, $w_0$ is the initial weights (bias) between input layer and $j$ neuron in the hidden layer and and $w_0$ is the initial weight of the output layer.$\varepsilon$ is the error of the approximation of the series at a given time. Sigmoid function, a continuous non-linear function, is the most commonly used activation function for neural network design with back propagation training.

This formula is used to predict the value of a data series $y$ at time $n$, $y(n)$, using the $p$ past values of the series. In the training phase, the true output is available and it was used as the input to the network as shown in Fig. 6.

During the training of the neural network aims to approximate the function $f$ by means of the optimization of the network weights and neuron bias. After training, the developed model is used to forecast the rainfall amounts.

The most common learning rule for the NAR network is the Levenberg–Marquardt backpropagation Algorithm. The Levenberg–Marquardt (LM) algorithm is widely used for training the NAR network it has fast convergence speed. The training time of the algorithm is decreased because LM algorithm uses an approximation of the Hessian matrix without direct calculation. The LM equation to update the weights is shown in (2).

$$w_{k+1} = w_k - [J^T J + \delta I]^{-1} J^T e(w_k) \quad (2)$$

Where $J$ is the Jacobian matrix. The Jacobian matrix has the first derivatives of the network error with respect the weights and biases. $I$ is the identity matrix. The variable $e$ is a vector of the network errors in every training sample. The Hessian matrix is $J^T J$ and $J^T e$ is the gradient. The parameter $\delta$ is the learning coefficient and is automatically updates based on the error at each iteration. The Levenberg–Marquardt update rule is a blend of both gradient descent and Gauss-Newton iteration. The effectiveness of the model is measured by using mean squared error (MSE) and R Squared in (3) and (4), respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (4)$$

Where $y_i$ is the actual value of the sample $i$ and $n$ is the number of samples,$\hat{y}_i$ is the forecasted value and $\bar{y}$ is the average value.

### III. Result and Discussions

The NAR has been trained using the historical rainfall data over the period 1938-2018 years. Accordingly, the historical data comprising input-target patterns have been divided into three parts: 70% for training, 15% for validation, and 15% as test data. The initial weights in the network are assigned randomly and they were adjusted at each iteration (i.e., epoch) to reduce the error. The procedure continued until the network output met the stopping criteria. The delay of input $n = 7$. NAR consists of one input layer, one hidden layer, and one output layer. The NARs are based on LM back-propagation training algorithms that were used for long-term time series prediction. MATLAB® software is used to build the models.
The NAR structure with 30 hidden neurons was found to be the most effective. It must be noted that increasing the number of neurons in the hidden layer makes a system more complex and decreasing the amount of neurons in the hidden layer will lower the computing power and generalization of the ANN. The results show clearer trends using seasonal year data sets in comparison with calendar year data sets as shown in Fig. 7 to Fig. 10 for two locations Dana and Muhai.

Both of prediction sets has very good accuracy where the value of mean square error (MSE) between the actual response and the target for test data is less than $5 \times 10^{-4}$ in all cases (Table III). The MSE error obtained can be considered negligible for nonlinear systems and it is less than what achieve in some other studies which enhance the credibility of rainfall prediction of the actual study. The annual prediction results show that using 25 neurons in hidden layer, 5 delay and 12 epoch produces predictions that perfectly match the input data (Table IV). This table present the prediction results of NAR NN with $n=25$ neurons in hidden layer and changing the number of delay.

Table IV show that $R=1$ when delay equal 5, so NAR performs best when delay is set 5. This means 5 earlier years’ prediction is meaningful to future prediction. When the delay is larger than 5, It will cause over fit issue which make the trained network less adaptable.

### Table III. Mean Square Error (MSE) Between the Actual Response and the Target for Test Data

| MSE   | Calendar Year | Seasonal Year |
|-------|---------------|---------------|
| CA0006| 4.32$\times 10^{-4}$ | 4.67$\times 10^{-4}$ |
| DB0002| 1.83$\times 10^{-5}$ | 4.92$\times 10^{-6}$ |
| DC0001| 1.42$\times 10^{-4}$ | 1.09$\times 10^{-4}$ |
| DC0002| 3.34$\times 10^{-3}$ | 2.12$\times 10^{-3}$ |
| DE0001| 1.34$\times 10^{-4}$ | 2.14$\times 10^{-4}$ |

### Table IV. Number of Delays and Epoch and R of Dana Location

| Delay | Epoch | R  |
|-------|-------|----|
| 1     | 22    | 0.32 |
| 3     | 15    | 0.93 |
| 5     | 12    | 1   |
| 8     | 43    | 0.8 |
| 10    | 91    | 0.63 |
| 12    | 31    | 0.78 |
| 15    | 75    | 0.48 |
Prediction results of NAR NN with delay=5 and changing the Number of neuron in hidden layer of Dana location is shown in Table V. In this table it is obvious that NAR performs best with number of neuron in hidden layer is set 25. This means 25 hidden is meaningful to future prediction. When number of neurons is larger, the NAR become worse.

Some models with different delays and hidden neurons combinations are shown in Table VI. This table shows that the model with 5 delay and 25 hidden neurons offers the best prediction accuracy where the value of Training R, Testing R and Validation R equal 0.98 or more.

The prediction in the time spam where the data is available is to validate the method and to check the model accuracy, which is an important issue. However, expanding the adapted method to cover certain time in the future is the real purpose of prediction. In this study, the future forecasting cover ten years beyond the last available rainfall data.

TABLE V. NUMBER OF NEURON IN HIDDEN LAYER AND EPOCH AND R OF DANA LOCATION

| Number of neuron in hidden layer | Epoch | R   |
|----------------------------------|-------|-----|
| 3                                | 55    | 0.25|
| 5                                | 40    | 0.33|
| 9                                | 81    | 0.45|
| 15                               | 14    | 0.64|
| 20                               | 22    | 0.78|
| 25                               | 12    | 0.98|
| 30                               | 27    | 0.89|

TABLE VI. RESULT OF SOME MODEL WITH DIFFERENT DELAYS AND HIDDEN NEURONS COMBINATIONS

| Model Structure                  | Training R | Testing R | Validation R |
|----------------------------------|------------|-----------|--------------|
| 3 delay, 5 hidden neurons        | 0.34       | 0.22      | 0.1          |
| 8 delay, 15 hidden neurons       | 0.54       | 0.31      | 0.2          |
| 2 delay, 20 hidden neurons       | 0.78       | 0.62      | 0.44         |
| 5 delay, 25 hidden neurons       | 1.0        | 0.98      | 0.99         |
| 10 delay, 30 hidden neurons      | 0.88       | 0.62      | 0.48         |

Fig. 11 to Fig. 14 show the plot results of predicting for 10 years of two locations for the best-achieved model. The red colored line represents the predicted rainfall values for 10 years.
The data presented in the Fig. 11 to Fig. 14 are summarize in Table VII. This table list long term recorded rainfall average and the rainfall average of the ten predicted years for each location. As the long-term average is almost the same in case of seasonal or calendar year base, the table prepared using the calendar year rainfall data. The comparison of average ten years predicted rainfall with the long term average revile that for Tafila station the predicted value is much higher than long term average (371 vs 286). For Muhi and Busira stations, the predicted values are quiet higher than the long-term average (149 vs 141 and 238 vs 234 respectively). For Rashadia station, the predicted rainfall is less than the long-term average (203 vs 218). While for Dana station the ten years average prediction is about 20% less than long-term rainfall average (286 vs 236). The difference between the long-term average rainfall and the average of predicted rainfall for the coming ten years seem to have a spatial pattern.

TABLE VII. LONG TERM RAINFALL AVERAGE AND AVERAGE OF 10 YEARS PREDICTED RAINFALL VALUES FOR THE FIVE STATIONS (MM / YEAR)

| Station | Muhi | Tafila | Busira | Rashadia | Dana |
|---------|------|--------|--------|-----------|------|
| Long term average | 142 | 286 | 234 | 218 | 286 |
| Predicted 10 years average | 149 | 371 | 238 | 203 | 230 |

In Tafila station, which locate in the center of study area the prediction, expect high increase (Fig. 15). In the next stations toward north and south (Muhi and Busira) the prediction is some increase. Moving more toward south to Rashadia station, the expectation is decrease. While in Dana station which is the most southern station, the expectation is high decrease.

This study showed that there is no clear trend for future expectactions, as it is evident from the above that some stations are expected to receive more precipitation, some less and some of them are almost stable. This result is consistent with the findings of [49] and [17], that there is no specific trend in their study areas. With regard to spatial related changes, many studies agree with this research, where they expected different changes in different places, but many of these studies used short time coverage data, and since precipitation rates in Jordan in general are highly variable, changing the length of the record may lead to different results [18].

Fig. 15. Average 10 Year Predicted Rainfall – Long Term Average for the Five Stations in the Study Area (mm).

IV. CONCLUSIONS

The results of the study show that the rainfall forecast efficiency is nearly identical for both seasonal and calendar years. The ANN technique proves to be highly efficient and accurate, and the MSE values were less than 5 * 10^-4 in all cases. The best accuracy was achieved with 25 hidden neurons and 5 delays. The results obtained showed that the future projection of precipitation is not uniform in the whole region. The expected rainfall amounts showed relative stability in two stations, a low decrease in one station, a significant decrease in one station, and a high increase in one station. According to the long-term mean, the projected shift in rainfall shows a spatial pattern, as there is a very large increase in the central station (Tafila) and the highest projected decline in the southern station (Rashadiya).

REFERENCES

[1] Keeling, R.F.; Piper, S.C.; Bollenbacher, A.F.; Walker, J.S. Atmospheric CO2 records from sites in the SIO air sampling network. In Trends: A Compendium of Data on Global Change; Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy; Oak Ridge, TN, USA, 2009.

[2] Hansen J, Sato M, Ruedy R, Lo K, Lea DW, Medina-Elizade M., Global temperature change. PNAS, vol. 103, no.39, 2006.

[3] Pendergrass AG, Knutti R, Lehner F, Deser C, Sanderson BM. Precipitation variability increases in a warmer climate. Sci. Rep.; vol. 7, 2017.

[4] Rowhani P, Lobell DB, Linderman M, Ramankutty NClimate variability and crop production in Tanzania. Agric. For. Meteorol, 2010.

[5] Shively GEInfrastructure mitigates the sensitivity of child growth to local agriculture and rainfall in Nepal and Uganda. Proc. Natl., 2017.

[6] Ministry Of Environment –MOE:-The National Climate Change Policy of the Hashemite Kingdom of Jordan 2013-2020: Sector Strategic Guidance Framework., 2013.

[7] Shehadah, N., The Climate in Jordan in the Past and Present. In. Hadidi, A. (Editor). Studies in the History and Archaeology of Jordan, II. Published by the Department of Antiquities. Amman, Jordan., 1985.

[8] Mountfort, G., Portrait of A Desert, The Story of an Expedition to Jordan. Collins. St James.s Palace, London, 1965.

[9] Nelson, J.B., Azraq Desert Oasis. Allen Lane. London, Fakenham and Reading., 1973.

[10] Alawneh, F., Balaawi, F., Hadad, N., and Shawabkeh, Y., “Analytical identification and conservation issues of painted plaster from Qaser Amra in Jordan,” International Journal of Conservation Science. vol. 2, no. 4, 2011.

[11] Nelson, J.B., Azraq . A Case Study. In. Hadidi, A, “ Studies in the History and Archaeology of Jordan, II” Published by the Department of Antiquities. Amman, Jordan., 1985.

[12] Alpert, A., Shafir, H., and Issahary, D., “Recent Changes in the Climate At the Dead Sea – a Preliminary Study,” Climate Change, vol. 37, no. 3, 1997.

[13] Abu Ghazleha,S., Kempea,S., Hartmannb, J., and Jansenb, N.,” Rapidly Shrinking Dead Sea Urgently Needs Infusion of 0.9 km3/a from Planned Red-Sea Channel: Implication for Renewable Energy and Sustainable Development. Jordan,” Journal of Mechanical and Industrial Engineering, vol.4, no. 1, 2010.

[14] Al Eisawi, D.M.H., “Water scarcity in relation to food security and sustainable use of biodiversity in Jordan,” Food security under water scarcity in the Middle East:Problems and solutions, vol 2, pp. 239 -248 , 2005.

[15] Ramirez, O.A , Ward, F. A., Al-Tabini, R. and Phillips, R., Efficient water conservation inagriculture for growing urban water demands in Jordan, Water Policy Vol 13 No 1 pp 102–124. (2011):

[16] Sharadqah, S., “Climate Change Trends in Tafila Governorate (Central West Jordan) in the Period 1938- 2006,” Journal of Natural Sciences Research. vol. 4, no.10, p 23-35, 2014.
