American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-03, Issue-11, pp-38-45
www.ajer.org

Research Paper Open Access

Development of Mathematical Models To Forecasting The Monthly Precipitation

Mustapha BEN EL HOUARI[1, 2], Omar ZEGAOUI[1], Abdelaziz ABDALLAOUI[2*]

[1)] *University Moulay Ismail, Faculty of Science, Department of Chemistry, Research Team "Applied Materials and Catalysis", BP 11201, Zitoune, Meknes, Morocco.*

[2)] *University Moulay Ismail, Faculty of Science, Department of Chemistry, Research Team "Analytical Chemistry and Environment", BP 11201, Zitoune, Meknes, Morocco.*

** Corresponding author: a.abdallaoui@gmail.com*

ABSTRACT: *The prediction of the precipitation is very important and challenging task in the modern world. In this study, the Multi Layer Perceptron (MLP) artificial neural network model was used for predicting the monthly precipitation in the region of Meknes in Morocco. Eight meteorological parameters (average temperature, maximum temperature, minimum temperature, pressure, moisture, visibility, wind speed and maximum wind speed) were used as predictors. The database used cover 209 months, from 1996 to 2013. The neural network architecture developed is [8-3-1]. The efficiency of the model measured by the calculation of the Root Mean Square Error and the linear correlation coefficient suggest that the studied meteorological parameters were related to the precipitation by nonlinear relationships. They also show that the ability of MLP neural network model was better than MLR, PLS and NLMR models.*

KEYWORDS**:** *MLP artificial neural network, MLR, NLMR, PLS, Precipitation, Prediction.*

I. INTRODUCTION

In the present-day context of climate change, precipitation has become one of the most important meteorological parameter. In semi-arid countries like Morocco, the prediction of this variable is primordial for different sectors such as agriculture and water resources management. In the last years, many studies have been conducted to rainfall forecasts in many countries such as Brazil [1], Mexico [2], or Turkey [3]. In general, theses authors have used Artificial Neural Networks (ANN) technique to predict the precipitation parameter. Indeed, artificial neural networks have been successfully used in various domains of science and engineering because of its ability to model both linear and non-linear systems without the need to make assumptions as are implicit in conventional statistical approaches. The ANN predictive technique has been used in weather events [4, 5], stock market [6]**,** particle interactions in high-energy physics [7], cloud classification and identification [8, 9], etc. Many models of ANNs have been developed for precipitation prediction such as Multi Layer Feed forward Neural Network (MLFN) [10]**,** Multi Layer Feed-forward Perceptron (MLFP) neural network [1]. However, for precipitation prediction, the Multi Layer Perceptron (MLP) neural network is the most widely used. In this paper, MLP neural network have been used to obtain a prediction model for the monthly precipitation in the region of Meknes in Morocco. The ANN analyses and their results are discussed and compared with the results obtained by using different traditional statistical models such as Multiple Linear Regression (MLR), Partial Least Squares regression (PLS) and Non-linear Multiple regression (NLMR). The meteorological parameters used as predictors for forecasting purposes are: average temperature (T), maximum temperature (T_{max}), minimum temperature (T_{min}), Pressure (Pr), moisture (H), Visibility (VV), wind speed (V) and maximum wind speed (V_{max}). The choice of these data is based on two criteria, namely the availability of data and the increased correlation between these parameters as demonstrated by others studies [11]**.** The objective of this study was to estimated daily precipitation from limited meteorological data using ANN model. The analyses and their results are discussed in this paper and compared with the results obtained by using the MLR, PLS, NLMR models. Finally we present the results obtained and the concluding remarks.

II. MATERIALS AND METHOD

Study Region and Data:

The study area comprises the region of Meknes (Figure 1). It is located in the central part of Morocco on plain of the Saïs with 33° 57’ Northern latitude, 05° 33’ Western longitude and 500 meter elevation. This region, with an area of approximately 4,560 km^2 is considered to be one of the best agricultural regions in Morocco.

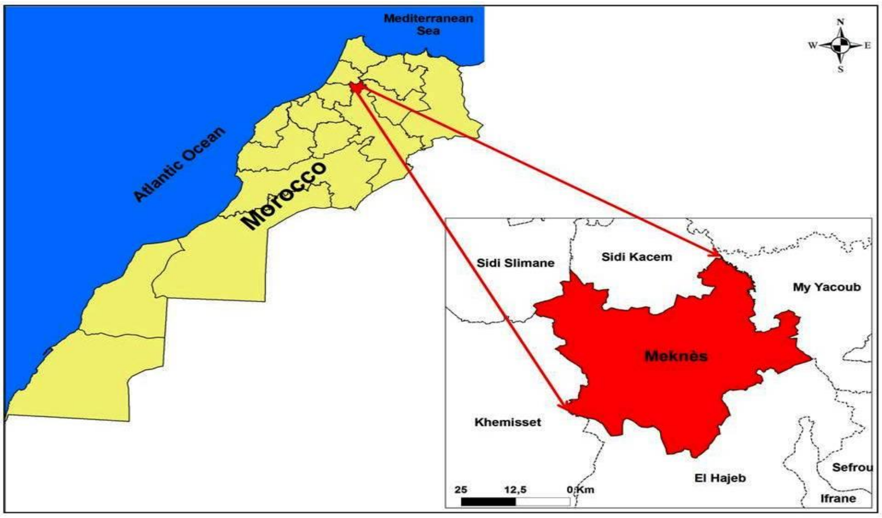

Figure 1: Geographic location of the region of Meknes

The average annual rainfall is close to 501.29 mm. The database used in this work cover 209 months, from 1996 to 2013. It contains the monthly precipitation (dependant variable: to explain) in addition to eight other meteorological variables (**T**: Average temperature**,** $\mathbf{T_{max:}}$ Maximum temperature, $\mathbf{T_{min:}}$ Minimum temperature**,** **Pr**: Pressure**,** **H**: moisture**,** **VV**: Visibility**,** **V**: Average wind speed and $\mathbf{V_{max}}$: Maximum wind speed**)**, chosen as independent variables (explanatory variables). Figure 2 show the evolution of the precipitation during the studied period. Summer is generally dry; the period of the largest rain lasts from October to May, with the number of days with monthly rain of 7 to 10 days. The average number of days with rain annually is estimated at 70 days and the annual total rainfall is around 501.29 mm.

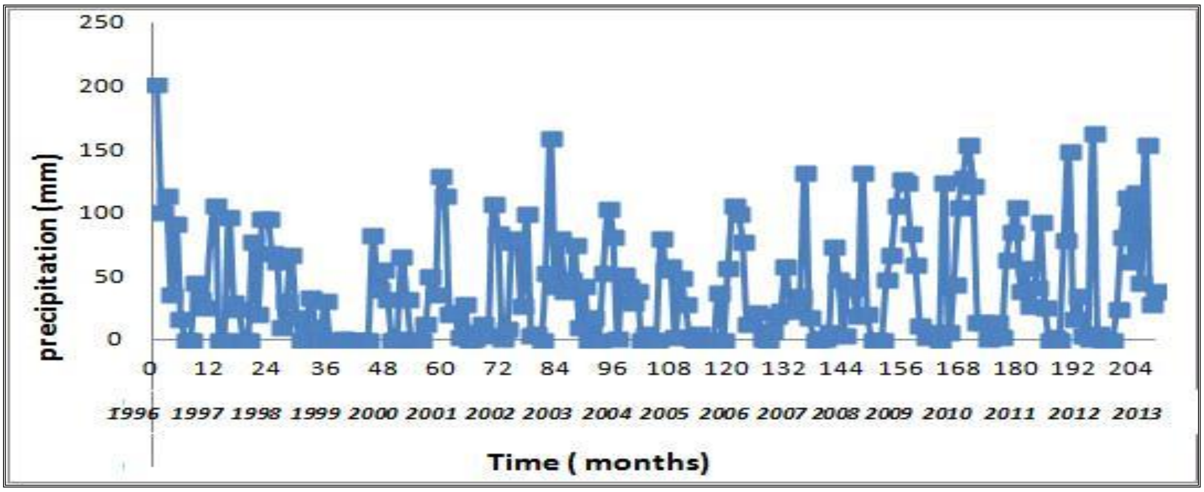

Figure 2: Evolution of the observed monthly precipitation in the region of Meknes between 1996 and 2013.

Performance evaluation of predictive model:

Two criteria are used to measure the model efficiency:

The Root Mean Square Error (RMSE) that represents the discrepancy between the computed and observed values is defined as: $RMSE = [\frac{1}{n}\sum_1^n(\acute{Y}_i - Y_i)^2]^{1/2}$ **(1)**

where **n** is the data number, $\acute{Y}_I$ are the predicted rainfall values, Y_i are the observed rainfall values

The linear correlation coefficient (R) is determined by: $R = \frac{COV(\acute{Y},\ Y)}{\sigma_{\acute{Y}}.\sigma_Y}$ **(2)**

Where *COV* is the Covariance and σ the average deviation,

In analyzing the different predictive models, the best answers for the model, RMSE go to zero and R go to one.

III. METHODOLOGY

Multiple linear regression (MLR):

The MLR model is used to predict values of a dependent variable from explanatory or independent variables. It is used to find the best linear model to predict the dependent value that produces the minimum error.

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n + e \quad \textbf{(3)}$$

Y is the dependant variable, a_0 is the regression constant and a_i (i = 1, 2, …n) are the regression coefficients and e is the error term.

Partial least squares regression (PLS):

Partial least squares (PLS) is a method for constructing predictive models when the factors are many and highly collinear. Note that the emphasis is on predicting the responses and not necessarily on trying to understand the underlying relationship between the variables.

Such as linear regression, PLS regressions equation is:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + e \quad \textbf{(4)}$$

Y is the dependant variable, b_0 is the regression constant and b_i (i = 1, 2, … n) are the regression coefficients and e is the error term.

Non-linear multiple regression (NLMR):

The purpose of the non-linear regression is to adjust the values of variables in the model to find the best curve which predicts Y in function of X. It can be expressed simply as:

$$Y_i = f(X_i, \theta) + \varepsilon_i \qquad i = 1,2, \dots n \quad \textbf{(5)}$$

Where Yi are the answers, f a nonlinear function depending on the vector $X_i = (X_{i1}, \dots X_{ik})$ and the parameter $\theta = (\theta_1, \dots \theta_p)$. ε_i is the residue.

Artificial Neural Network (ANN):

An artificial neural network is an interconnected group of artificial neurons that has a natural property for storing experiential knowledge and making it available for use. ANN consists of a large number of processing elements called neurons, which are arranged in different layers in the network: an input layer, an output layer and one or more hidden layers [12].

The basic ANN architecture is shown in Figure 3

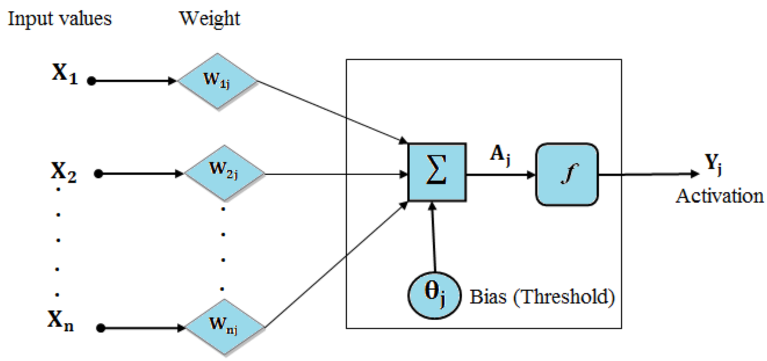

Figure 3: Schematic presentation of neural network [13].

The neuron receives input signals, processes them and sends an output signal [14]. Each neuron is connected with at least one other neuron and each connection is represented by a real number called a weight. The weights are adjusted iteratively so that the network attempts to produce the desired output [15]. Mathematically, the process can be expressed following the equation (6) [14].

$$Y_k = f[\sum_{k=1}^{n} W_k X_k + b_k] \quad \textbf{(6)}$$

Where W_k represents the synaptic weight, X_k is the input value (k = 1, 2..., n), b_k is the bias of neuron, f is the transfer function, and Y_k is the output. In this work, the sigmoid activation function has used in hidden layer [1]. It defined for any variable X as:

$$f(X) = \frac{1}{1+\exp(-X)} \quad \textbf{(7)}$$

For the output layer, a linear transfer function has used [8].

$$a = f(n) = n \quad \textbf{(8)}$$

Among the 209 data set, 169 was chosen randomly for the training phase and the rest (40 data) for the test phase. Precisely 20%, this is the best distribution of the test phase. The first option is connected to the distribution of the database in two bases: training phase and test phase based on the calculation of the root mean square error. RMSE in Table 1.

Table 1: Root mean square error of training and testing phase for different distributions of the data.

	RMSE Training phase	RMSE Testing phase
Training phase : 90% Testing phase : 10%	0.22	0.20
Training phase : 80% Testing phase : 20%	**0.19**	**0.18**
Training phase : 70% Testing phase : 30%	0.21	0.20

A study has been conducted to choice the number of hidden neurons. The RMSE results of the integrated ANN obtained for different hidden layer are presented in table 2.

Table 2: Root mean square error of training and testing phase for different Number of neuron of the hidden layer.

Number of the neuron of hidden layer	2	**3**	4	5	6	7
RMSE Training phase	0.23	**0.19**	0.24	0.29	0.20	0.29
RMSE Testing phase	0.20	**0.18**	0.20	0.21	0.20	0.19

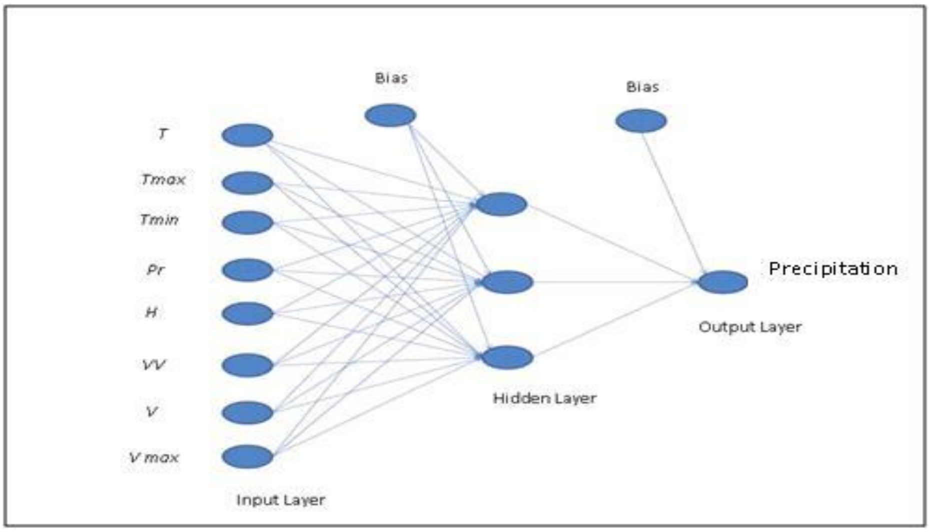

Figure 4: The developed neural network architecture [8-3-1].

An optimal ANN architecture may be considered as the one yielding the best performance in terms of error minimization. According to the results by increasing the number of neuron of the hidden layer, the training and testing RMSE values show that the best structure obtained is with 3 neurons of the hidden neurons corresponding to the lowest RMSE values. So the developed neural network architecture is [8-3-1], as presented in figure 4.

The model used in this study is the type of multilayer perceptron (MLP) artificial neural network with a back propagation of the square error. The rule of the back-propagation is to propagate the errors through the network and allow the adaptation of the hidden units. The Levenberg-Marquardt is used to optimize the weights of the network with transfer functions: sigmoid and linear.

IV. RESULTS AND DISCUSSION

Statistical models (MLR, PLS, NLMR)

We obtained the equations (9), (10) and (11) respectively after analyze by the MLR, PLS and NLMR:

[MLR] :
$$Y_{MLR} = 2766.61 - [4.11 \times T] - [13.50 \times T_{max}] + [16.37 \times T_{min}] - [2.39 \times Pr] - [0.91 \times H] - [9.34 \times VV] - [0.09 \times V] + [2.89 \times V_{max}] \quad \textbf{(9)}$$

[PLS] :
$$Y_{PLS} = 2149.23 - [4.10 \times T] - [12.91 \times T_{max}] + [16.29 \times T_{min}] - [1.96 \times Pr] + [2.19\ 10^{-2} \times H] + [3.54 \times VV] - [3.31 \times V] + [4.36 \times V_{max}] \quad \textbf{(10)}$$

[NLMR]:
$$Y_{NLMR} = 37423 + [16.6 \times T] - [44.6 \times T_{max}] + [27.2 \times T_{min}] - [732.94 \times Pr] - [1.37 \times H] - [48.14 \times VV] - [1.16 \times V] + [14.97 \times V_{max}] - [0.40 \times T^2] + [0.54 \times T^2_{max}] - [0.55 \times T^2_{min}] + [0.35 \times Pr^2] + [5.69\ 10^{-3} \times H^2] + [3.05 \times VV^2] + [3.73\ 10^{-2} \times V^2] - [0.30 \times V^2_{max}] \quad \textbf{(11)}$$

The signs of the coefficients for the variables in the model (MLR) are similar to the PLS model with the exception of the moisture H and Visibility VV. This compatibility signs shows the probable existence of a strong correlation between the two models.

The correlation coefficients obtained by statistical models (MLR, PLS, NLMR) for training series are respectively 0.68, 0.80 and 0.82. In fact, the NLMR method is most effective, the correlation coefficient between the observed and predicted monthly precipitation is significantly important (R = 0.82) compared to the MLR and PLS methods.

Artificial Neural Network (ANN):

In this study, ANN model has been developed with 8 neurons in the input layer (**T**, $\mathbf{T_{max}}$, $\mathbf{T_{min}}$, **Pr**, **H**, **VV**, **V** and $\mathbf{V_{max}}$), three neurons in the hidden layer and one neuron in the output layer for the prediction of precipitation. Table 3 shows the regression coefficient and the root mean square error of training and testing phases obtained by using the four prediction models. It is clear that the developed ANN model shows the best correlation coefficient (R=0.97 and 0.98 for training and testing phases respectively) corresponding to a RMSE = 0.19 and 0.18 respectively while the MLR calculation lead to a wrong correlation value (R=0.68 and 0.81 for training and testing phases respectively) corresponding to a RMSE = 3.7 and 3.5 respectively.

Table 3: Error criteria during training and testing process of different model.

Model	Training phase		Testing phase	
	R	RMSE	R	RMSE
MLR	0.68	3.7	0.81	3.5
PLS	0.80	2.6	0.84	2.4
NLMR	0.82	2.5	0.85	2.3
ANN	**0.97**	**0.19**	**0.98**	**0.18**

The Figure 5 shows the relationships and coefficients of regression between the observed and the predicted precipitation values using the MLR, PLS, NLMR and ANN models. The figures 5A, 5B and 5C indicate clearly that the points obtained by using MLR, PLS and NMLR prediction models are not uniformly scattered around the regression lines. However, figure 5D relative to ANN prediction model shows a good correlation between observed and predicted rainfall values. These results and those presented in table 3 suggest that the studied meteorological parameters are related to the precipitation by nonlinear relationships. Also, these results show that the ability of ANN model to predict the precipitation values are better than MLR, PLS and NLMR models, the ability of ANN model to predict the mid-range values was better than the ability of MLR model as shown in ANN diagram, the points are scattered closer to the straight line than the clusters in MLR, PLS and NLMR diagrams (Figure 5).

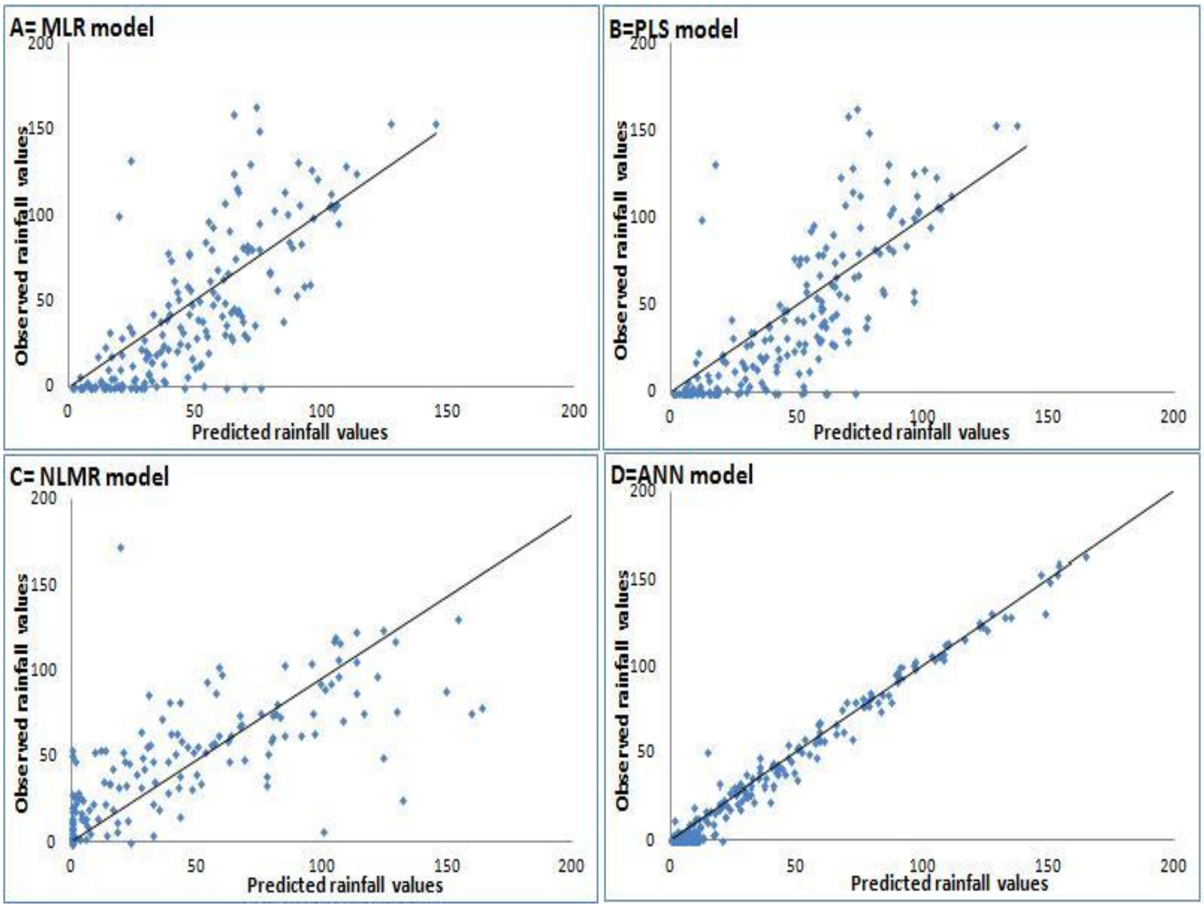

Figure 5: Comparison between the observed total precipitation data and the results obtained from (A): MLR, (B): PLS, (C): NLMR and (D): ANN models

The residue is the error committed by the models established by each individual method on a sample of model construction; Figure 6 shows the plots of the distribution of the residuals based on the simulated values obtained during the training and testing phases using the four predictive models. The figures 6A, 6B and 6C show that the points representing the residues are not uniformly scattered around the abscissa axis zero. However, the figure 6D shows that the points representing the residues fall on the abscissa axis zero, indicating that the more perfectly prediction model is the ANN in comparison with the MLR, PLS and NLMR prediction models.

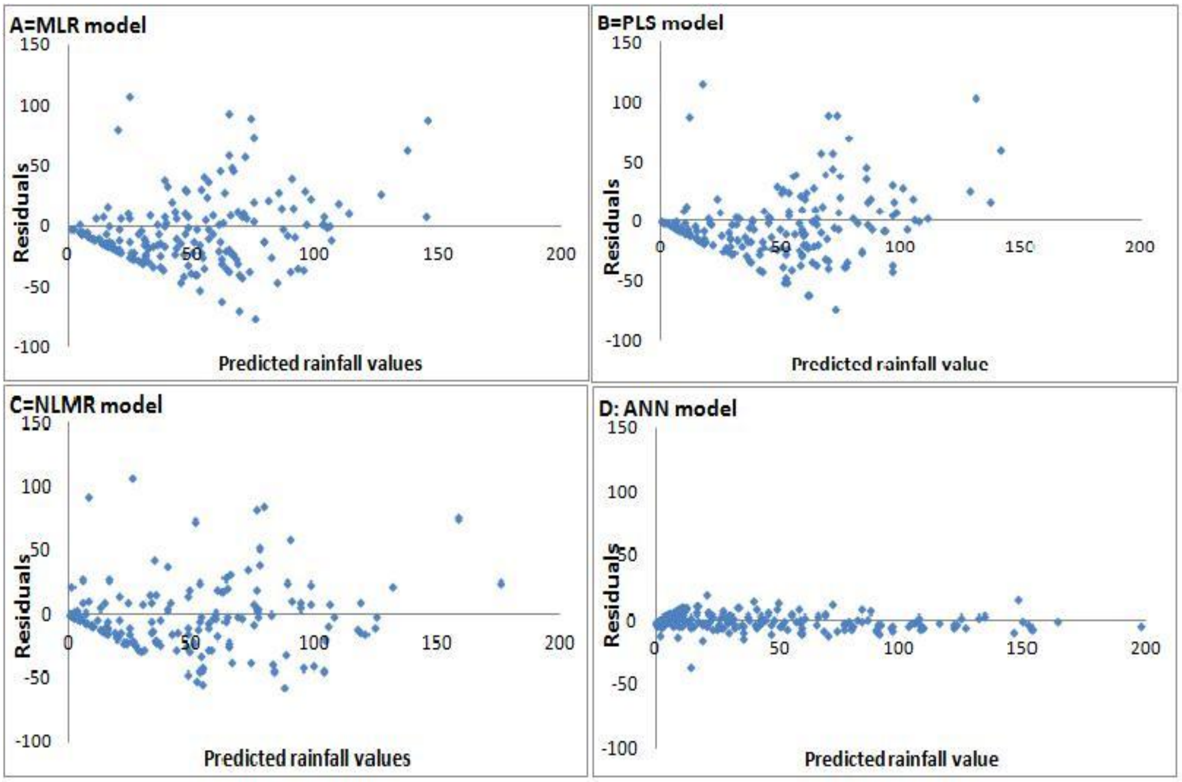

Figure 6: Distribution of residuals obtained by the four prediction models.

V. CONCLUSION

ANN model has been developed to predict the monthly precipitation parameter for the region of Meknes, Morocco. Based on the RMSE and R values, the appropriate architecture of the neural network was [8-3-1] and the ANN model appears capable of providing accurate predictions of the monthly precipitation by comparison with the conventional statistical models (MLR, PLS, NLMR). The results show a significant capacity for prediction for monthly precipitation contents with a linear correlation coefficient (R) of 97% and a low Root Mean Square Error (RMSE) of 0.19 for the database used. For multiple linear regression (MLR), Partial least squares regression (PLS) and Non-linear multiple regression (NLMR) the results are less significant with a linear correlation coefficient (respectively 0.68, 0.80 and 0.82). This shows that the parameters are associated with the monthly precipitation by a non-linear relationship. For the prediction of monthly precipitation, the use of a neural model configuration [8-3-1] gave better results.

REFERENCES

[1] M.C.V. Ramirez, N.J. Ferreira, and H.F. Velho, Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region, *Journal of Hydrology*, *301*, 2005, 146-162.

[2] C. Tereza, Downscaling large-scale circulation to local winter rainfall in north–eastern Mexico, *International Journal of Climatology*, *17*, 1997, 1069-1082.

[3] K. Saplioglu, M. Cimeny, and B. Akman, Daily precipitation prediction by artificial neural networks. Proceedings of the 4th International Scientific Conference on Water Observation and Information System for Decision Support Water Observation and Information System for Balkan Countries. Ohrid, Republic of Macedonia, 2010.

[4] C. Marzbam, and G. Stumpf, Multiresolution wavelet transform and neural networks methods for rainfall estimation from meteorological satellite and radar data. *J. App. Meteor.*, *35*, 1996, 617-626.

[5] K. Hsu, H.V. Gupta, X. Gao, and S. Sorooshian, Rainfall Estimation from Satellite Imagery, Chapter 11 of Artificial Neural Networks in Hydrology, *Edited by R.S. Govindaraju and A.R. Rao, Published by Kluwer Academic Publishers*, 2000, 209-234.

[6] E. Collins, S. Ghosh, and C. Scofield, An application of a multiple neural network learning system to emulation of mortgage underwriting judgements. *Proceedings of the IEEE International Conference on Neural Networks*, 1988, 429.

[7] R. Kantowski, and C. Marzban, A neural network for locating the primary vertex in a pixel detector, *Nucl. Instrum. Methods phys. Res*, *355*, 1995, 582-588.

[8] R.L. Bankert, Cloud classification of a advanced very high resolution radiometer imagery in maritime regions using a probabilistic neural network. *J. Appl. Meteor*, *33*, 1994, 909-918.

[9] J.E. Peak, and P.M. Tag, Sesmentation of satellite imagery using hierarchical thresholding and neural networks. *J. Appl. Meteor., 33*, 1994, 605-616.

[10] K.C. Luk, J. Ball, and A. Sharma, An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer modelling, 33(6)*, 2001, 683-693.

[11] H. AL-Naimi.Rasheed, M. AL-Salihi. Ali, I. Bakr. Dher, Neural network based global solar radiation estimation using limited meteorological data for Baghdad, Iraq. *International Journal Of Energy And Environment*, 2014, 79-84.

[12] A.J.P. Kumar, and D.J.K. Singh, 2008. Artificial neural network-based wear loss prediction for a 390 aluminium alloy, *J. of Theoretical and Applied Information Technology, 4(10)*, 2008, 961-964.

[13] H. El Badaoui, A. Abdallaoui, I. Manssouri, and H. Ousmana, The prediction of moisture through the use of neural networks MLP type. *Journal of Computer Engineering*, 2013, 66-74.

[14] S. Haykin, Neural Neural Network and Its Application in IR, *a comprehensive foundation, 2nd edition Upper Saddle Rever, New Jersey: Prentice Hall*, 1999, 842p.

[15] M. Safa, S. Samarasinghe, and M. Mohsen, Modelling fuel consumption in wheat production using neural networks. *18th World International Congress on Modelling and Simulation, Cairns, Australia, 13-17 July 2009*, 2009, 775-781.