Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

A fuzzy based model for rainfall prediction

Bilal Zahran[a*], Belal Ayyoub[b], Waleed Abu-Ain[c], Wael Hadi[d] and Sulieman Al-Hawary[e]

[a]*Department of Engineering & AI, Al-Salt Technical College, Al-Balqa Applied University, Jordan*
[b]*Department of Electrical Engineering, Faculty of Engineering Technology, Al-Balqa Applied University, Jordan*
[c]*Computer Science Department, Applied College, Taibah University, Saudi Arabia*
[d]*Department of Information Security, Faculty of information Technology, University of Petra, Jordan*
[e]*Department of Business Administration, Business School, Al al-Bayt University, Jordan*

C H R O N I C L E

Article history:
Received: May 20, 2022
Received in revised format: September 26, 2022
Accepted: November 30, 2022
Available online: December 2 2022

Keywords:
Fuzzy logic (FL)
Weather forecasting
Rainfall prediction

A B S T R A C T

Of all the current challenges faced by Jordan, the most severe is the inadequacy of the water supply. The country is almost entirely reliant on rainfall, whose pattern, however, is highly variable in terms of its frequency, regularity, and quantity. Evidently, therefore, the ability to anticipate rainfall accurately is critically important for the effective planning and management of water resources in Jordan, and particularly in agricultural areas. Influenced by a range of factors such as temperature, relative humidity, and wind speed, rainfall is a stochastic process. This paper suggests the use of a fuzzy model that draws upon data gathered at 26 stations situated in a range of locations throughout Jordan. The model is capable of forecasting seasonal rainfall relating to a specific station. Its ability to deliver predictions with an acceptable degree of accuracy has been demonstrated, and it can be concluded from this that the fuzzy technique can provide a model that is capable of efficiently forecasting seasonal rainfall.

1. Introduction

Among the services provided by meteorological offices globally, one of the most important and challenging is the forecasting of the weather—a complex process that integrates a wide range of specialist technological expertise. One of the fundamental elements involved in this is water, which is essential to human survival and necessary for a wide range of vitally important activities. These include the production of food through agriculture, which is wholly reliant on rainfall for its success, but rainfall is also critical for many other processes and activities that are critical to human life. The variables that determine climatic conditions—such as minimum and maximum temperatures, humidity, and rainfall—are in a constant state of flux over time. These generate a time series with respect to each parameter, and this may be utilized to create a prediction model, either statistically or via other means that draw upon the time series data. (Ozone layer, temperature, relative humidity, etc.) There is consequently a requirement for effective control over these variable factors in order to obtain accurate forecasts of rainfall, and a number of computational methods have been proposed with a view to accomplishing this. In the context of the adequacy of water resources, Jordan is considered to be one of the most deprived nations. Located within the eastern Mediterranean climate zone, the country is characterized by hot and dry summers and cool and wet winters. Its rainfall profile shows an irregular distribution across the various regions as well as substantial annual variation in terms of its quantity and timing (Jordanian Ministry of Water and Irrigation Publication, 2015; Raddad, 2005; Al-Ansari et al., 2014; Zahran, 2015; Janarthanan et al., 2021), with a rainy season extending from October to April. A significant proportion of Jordan's landmass is arid or semi-arid; here, the annual rainfall amounts, on average, to less than 200 millimeters, most of which is lost to

* Corresponding author.
E-mail address: zahranb@bau.edu.jo (B. Zahran)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2022.12.001

evaporation, whereas in the north-west region, a maximum of some 600 millimeters per annum may be received. The country is primarily dependent on rainfall for its water resources, but water availability is highly variable due to the dry climatic conditions. The severity of the water resources challenge may be observed by tracking the per capita supply, which by 1977 had risen to 429.9 m^3, compared with only 94 m^3 in 2018. This dramatic shift in consumption is accounted for by the rapid population increase and the lack of water resources (Raddad, 2005; Al-Ansari et al., 2014). Jordan's rainfall and the parameters that inform it are uncertain and non-linear. Jordan is therefore considered to be suitable for the fuzzy logic algorithm approach to forecasting, which is commonly supported in contemporary research in the field of rainfall prediction.

FL is a type of many-valued logic that employs approximate rather than precise reasoning. FL variables may possess a truth value varying between 0 and 1 (Zadeh, 1965; Logic et al., 1999; Kasabov, 1998), as distinct from typical binary sets, where the value of a variable is restricted to either “true” or "false." A FL consists of the nonlinear mapping of a set of input data to scalar output data and includes the four basic elements of a FL system: the fuzzifier, the rules, the inference engine, and the defuzzifier. At the design stage, the fuzzy logic process includes the following phases: definition of the linguistic variables and terms; construction of the membership functions; and construction of the rule base. In the inference phase, the stages include: employing the membership functions; converting crisp input data into fuzzy values; evaluating the rules within the rule base; combining the outcomes of each rule; and defuzzification, which involves converting the output data to real values (Logic et al., 1999; Kasabov, 1998). It is notable that a number of scientific domains that involve uncertainty have beneficially deployed fuzzy-based models.

This paper is structured as follows: Section 2 considers existing work in the fields of rainfall prediction and fuzzy-based systems. In Section 3, proposals for FL models are explored, while Section 4 provides reflections on the results. Finally, the conclusions of this research are presented in Section 5.

2. Literature review

The existing literature suggests that Artificial Intelligence (AI) algorithms are eminently suitable for application within the weather forecasting field, and in particular for the prediction of rainfall. Adopted widely within numerous fields characterized by nonlinear patterns, FL lends itself especially well to utilization in the development of models to predict weather parameters. This is due to the substantial levels of uncertainty inherent in weather forecasting, as attested by the current literature (Ramadoss et al., 2020; Rahman, 2020; Kumar, 2019; Agboola et al., 2013), in which a number of current studies have deployed FL for this purpose. A discussion of these studies is provided in this section. In Ramadoss et al. (2020), acceptable results were produced in a study that utilized an intelligent fuzzy model where temperature (T) and wind speed (SW) were used to forecast the rainfall (FR) rate. This involved applying the fuzzy rule, whereby the antecedent and consequent statements relate the input variable to each other in order to determine the results. The study presented in Rahman (2020) combined the input factors of temperature and wind speed with a single output variable—the amount of predictable rainfall. The graph's diagram is constructed using eight equations representing temperature and wind speed, which correspond to the membership values. Eight equations for distinct categories are also created by rainfall. The environmental conditions that increase rainfall incidence are represented by fuzzy levels. Membership functions are derived following the minimum composition of the inference section of the fuzzification undertaken for temperature and wind speed.

In Kumar, 2019, the researchers investigated the use of a FL method to examine rain prediction, given that this can assist in forecasting short-term load. The security analysis of generational short-term load forecasting is a highly valuable instrument for unit commitment. Based on the results, it was concluded that FL is a reasonably accurate technique for predicting the amount of rainfall when using wind speed and temperature data. As previously noted, the accurate prediction of rainfall plays an important role in enhancing the management of water resources. Effective connections between water authorities across the country and the accurate forecasting of rainfall are also critically important in facilitating the monitoring and avoidance of disaster conditions such as drought and flooding. The source (Agboola et al., 2013) explores whether FL may be utilized to model rainfall in south-western Nigeria. Here, the FL model consists of two functional components—on the one hand, the fuzzy reasoning or decision-making element, and on the other, the knowledge base. The authors computed the prediction accuracy and determined, according to the results obtained, that the fuzzy technique is indeed capable of efficiently managing the collected data. Their model displayed both flexibility and the capability to replicate a poorly defined connection between the input and output variables.

3. Methodology

3.1 Area of study (Jordan)

Located on the Asian continent between latitudes 29° and 34°N and longitudes 35° and 40°E, Jordan experiences a semi-dry climate in the summer with average temperatures in the mid–30s °C (approximately 86 °F) and a chilly winter, when the temperature is on average approximately 13 °C (55 °F). The rainy season extends from November to April, with December and February being the wettest of the winter months. Other than in the northwest of the country, which receives between 250 and 600 millimeters (10 and 18 inches) of precipitation annually, Jordan typically receives under 100 millimeters (4 inches) of rain. The present study separated Jordan into four distinct areas—the north, south, middle, and desert—with each area having rainfall and climatic parameters that are distinct from the others. Fig. 1 depicts the subdivision into areas. In defining the four areas in which stations were to be located, the sole criterion considered was seasonal rainfall between October and April.

Fig. 1. Map of Jordan with each of the four regions (areas of study) shown in different colors

3.2 Data set and data pre-processing

Monthly rainfall data was obtained from the national meteorological department of Jordan for the period 1977–2020. The data was obtained from 26 weather stations across Jordan. To address the issue of dirty, incomplete, and noisy real-world data, data cleaning was conducted, which also addressed duplicate records, data entry inaccuracies, and inconsistencies such as those arising from the use of multiple data sources. This involved the replacement with mean values of any values missing from the seasonal rainfall data. Outlier data—such as observations that were distant from others—were removed. A sample of the climate dataset for the middle region, including the capital, Amman, is provided in Table 2 and Fig. 2.

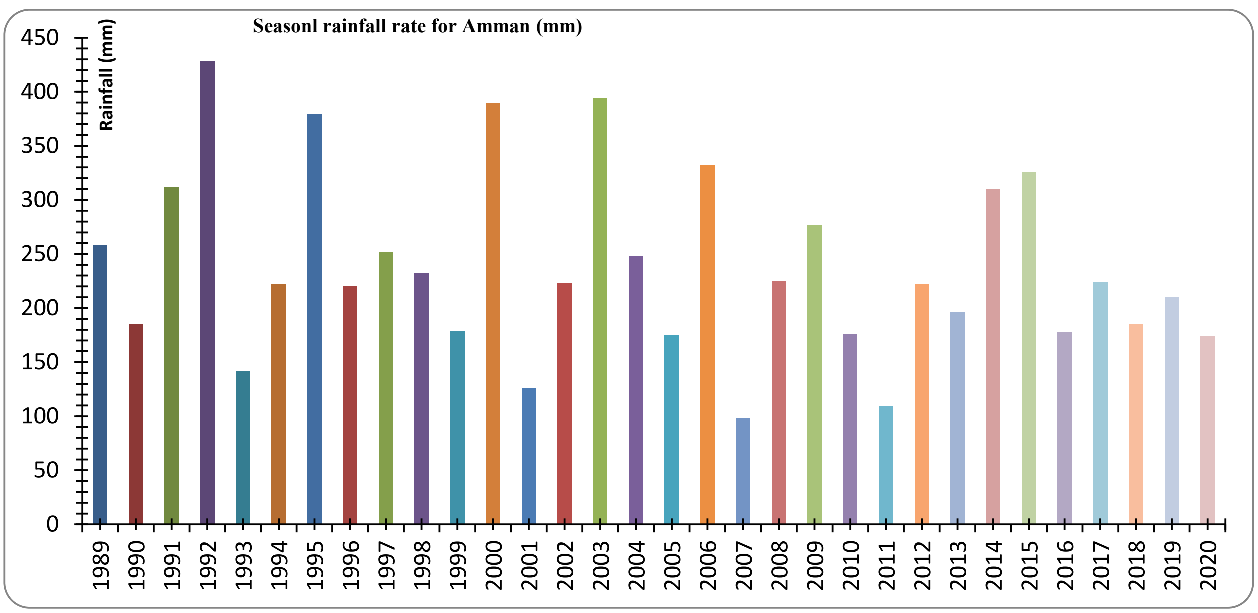


Fig. 2. The seasonal rainfall rate of Amman (Middle region stations) for period between 1989 and 2020

Table 1
Sample of dataset for middle region in Jordan

TMEAN	CLOUD	SWPEED	HUMDAYS	PREC
		The Middle		
9.7	**3.4**	**4.3**	**18**	**37**
14.7	**3.4**	**7.4**	**4**	**57.6**
6.9	**4.5**	**4.8**	**13**	**73.9**
1988	**8**	**3.9**	**6.8**	**15**
7.95	**5.54**	**8.8**	**19.9**	**140.1**
7.9	**5.9**	**8.3**	**23.8**	**173**

5.1	**5.7**	**9.8**	**22**	**199.8**

3.3 Fuzzy model

A fuzzy logic model may also be known as a fuzzy inference system. In the current study, the two functional elements of the model are: firstly, the knowledge base, consisting of a number of fuzzy "if-then" rules; and secondly, a database that describes the membership functions of the fuzzy sets used in the fuzzy rules. The inference operations based on these rules are performed by the fuzzy reasoning or decision-making unit, which is founded upon the knowledge base.
A typical process for building a fuzzy expert system will consist of the five steps set out below:

1. Define the linguistic variables and specify the problem;
2. Identify the fuzzy sets;
3. Create fuzzy rules by eliciting and constructing them;
4. Encode the fuzzy sets, fuzzy rules and fuzzy procedures in order to conduct fuzzy inference on the expert system;
5. Assess and adjust the system;

Steps 1 and 2: Defining the problem, variables and fuzzy sets

Applying steps 1 and 2 above to the current model, the involvement of several linguistic variables is noted. These are: temperature (T), wind speed (SW), cloud (C), humidity (H), and rainfall rate (FR). The linguistic values selected for the typical linguistic variables are: VL, L, N, H, and VH, the meanings of which are defined in Table 2. The mathematical approach to deriving the fuzzy set of a typical fuzzy variable is presented in Fig. 3.

Table 2
Fuzzy Variables meaning

Abbreviations	Meaning
VL	Very low
L	Low
N	Normal
H	High
VH	Very high

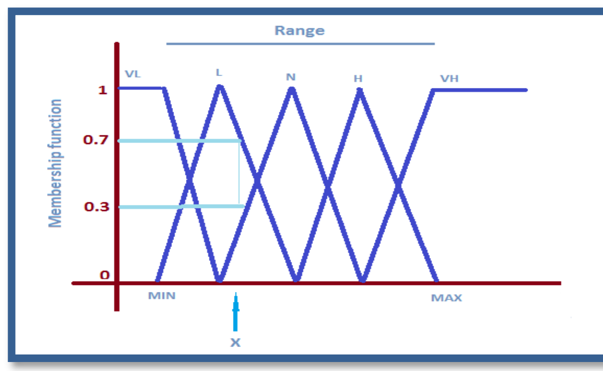


Fig. 3. Fuzzy sets of Membership function (μ) and corresponding fuzzy levels

In the calculation operation, the range of the fuzzy variables, between their minimum and maximum values, is separated into an ascending-order numerical scale, beginning with the minimum. Fig. 3 displays the range and fuzzy levels for any fuzzy set of objects in a triangle functional diagram, with the range divided into five equal sub-ranges, each representing a fuzzy level. They are displayed in ascending order and abbreviated as VL, L, N, H, and VH, with abbreviations and definitions of each level provided in Table 2.

Step 3: Constructing Fuzzy Rules

In this stage, appropriate production rules are constructed. These consist of the antecedent and consequent sections of the fuzzy rule, followed by algorithms using logic based on the previous experience of decision makers. The truth table that supports identification of the potential fuzzy rules can be found at Table 3.

Table3
Fuzzy RuleTruth Table

T \ SW	VL	L	N	H	VH
VL	VL	L	H	VH	VH
L	VL	L	L	N	H
N	VL	L	N[1]	N	L
H	VL	VL	VL	VL	VL
VH	VL	VL	VL	VL	VL

[1]The intersection of row (T) and column (SW) produce rainfall (FR) value. The rule of this value is : IF (TP is N) and (WS is H) then (RF is N)

Step 4: Coding the fuzzy system

In step 4, the design of the fuzzy sets and fuzzy rules involves coding the fuzzy system. In this study, the MATLAB Fuzzy Logic Toolbox was deployed to achieve this.

Step 5: Assessment and adjustment

In Step 5, the final phase, the system is assessed and adjusted accordingly. This is the most challenging stage, as it is essential in determining whether or not the fuzzy system fulfils the requirements set out for it.

3.4 Performance evaluation—error measures

A number of error metrics were utilized to establish the effectiveness of the fuzzy rule-based model, as set out below:

1. **Prediction Error (PE):**

$$PE = \frac{(|y\ predicted - y\ actual\ |)}{y\ actual} \tag{1}$$

where the PE is sufficiently minimal—defined as near to 0— the prediction model is considered good.

2. **Root Mean Square Error (RMSE):**

$$RMSE = \sqrt{\frac{\sum_j^N (y_j - \hat{y}_j)^2}{N}} \tag{2}$$

This metric is typically used to determine the extent of any divergence between the prediction provided by the model and the actuality of the entity being modeled.

3. Mean Absolute Error (MAE):

$$MAE = \frac{|y_j - \hat{y}_j|}{N} \tag{3}$$

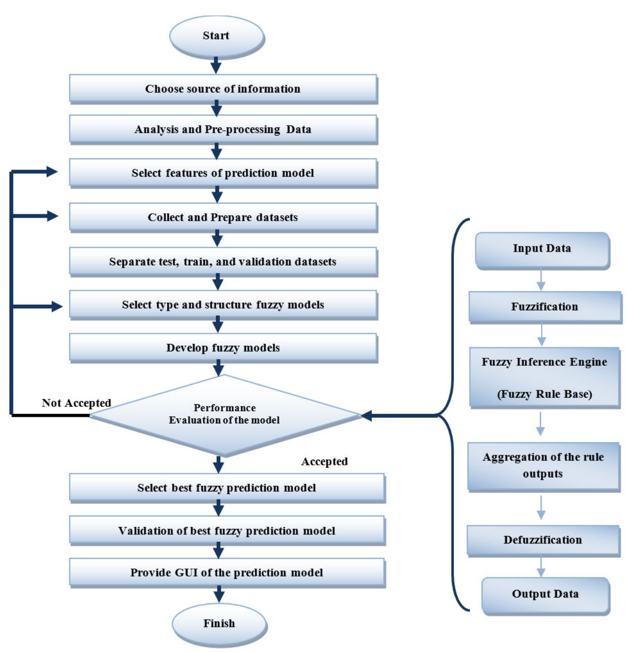
The lower the MAE generated by this calculation, the more accurate the model.

4. Accuracy:

$$Accuracy = 100 - RMSE \tag{4}$$

where y_j and $\hat{y}_j$ are observed and predicted values for rainfall, respectively and N is the number of observations.

Fig. 4 shows the fuzzy logic Modeling Procedures and performance evaluation.



A. Modeling Procedures B. Performance Evaluation

Fig. 4. Fuzzy Logic Modeling Procedures: A. Modeling Procedures, B. Performance Evaluation

3.5 Primary proposed fuzzy model

The scheme outlined below was utilized to construct a primary model, which was then successively improved until the final version, having the lowest possible error rate, was achieved. The primary model had two input variables: wind speed (SW) and temperature at a specific time (T), as well as one output variable: estimated rainfall. The input variables were selected on the basis of their proven influence upon the occurrence of rainfall. The values of the input variables are grouped into fuzzy levels using the linguistic variable. With a membership value, the input variable will belong to one or, at most, two of these levels. An input value is translated to its corresponding membership function (MF) value by establishing parameter types, ranges, and rules to identify all the membership values for any specific input variable. Further detail of the primary model is depicted in Fig. 5 and Table 4.

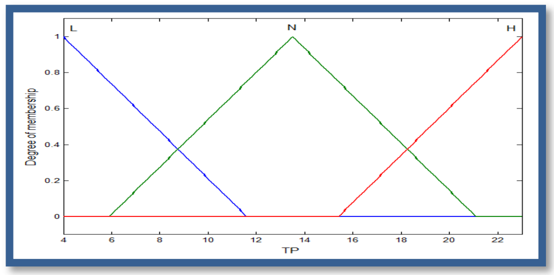


Fig. 5. Fuzzy set and membership functions related to temperature (T) – primary model

Table 4
Sample of primary model Rules

1.	If (SW is L) and (T is L) then (FR is L)
2.	If (SW is L) and (T is N) then (FR is L)
3.	If (SW is L) and (T is H) then (FR is L)
4.	If (SW is N) and (T is L) then (FR is L)
5.	If (SW is N) and (T is N) then (FR is N)

3.5.1 Results of primary model and discussions

Fig. 6 presents the preliminary results of our survey

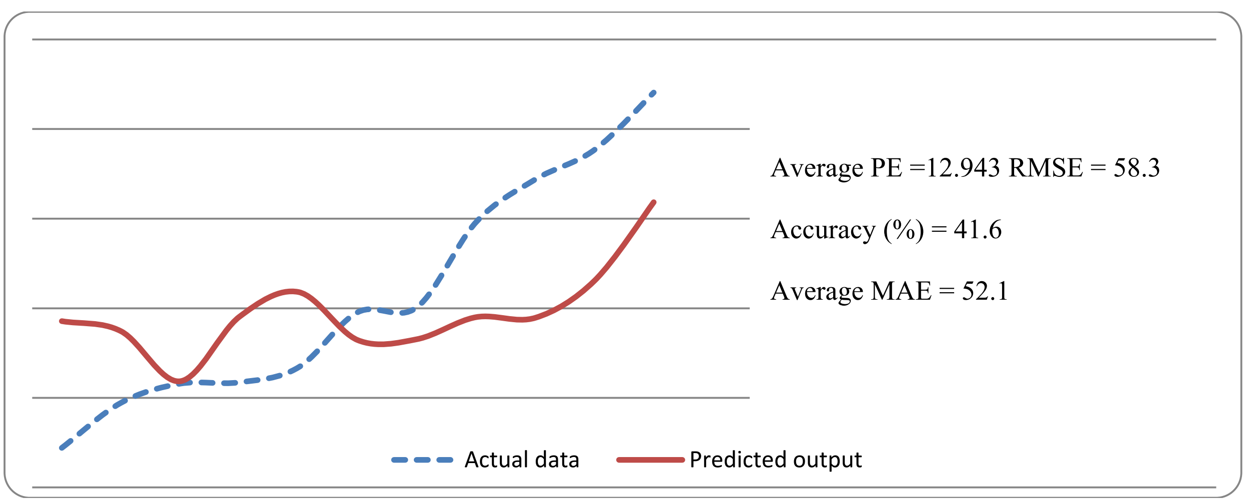


Fig. 6. Actual values vs. predicted values (Primary model –Middle region)

3.5.2 Results of primary model and discussions

The results demonstrated that the primary model is not accurate, achieving only a 41.6% measure of accuracy, which is unsatisfactory. In a comparison of the i^{th} and $(i-1)^{th}$ seasons utilizing real data and total results, it was established that the value of the computed FR diverges from the corresponding pattern of decrease or growth with identical ranges of SW and T data.

This is particularly the case where the actual value of FR is situated within the upper range. On the basis of this assessment, it was concluded that substantial improvement of the primary model was required.

3.5.2 Improvement of the primary model

The fundamental issue with the previous initial fuzzy-based time series forecasting model was its inability to successfully predict the rainfall. A review of the several procedures carried out to improve the model is described here:

The fundamental issue with the previous initial fuzzy-based time series forecasting model was its inability to Handling the missing and anomaly values found in the dataset with appropriate means.

- Choosing the appropriate order (past values) of time-series data experimentally.
- Reconstructing the fuzzy sets according to the importance of the climatic parameter to rainfall prediction
- Rebuilding the fuzzy rules carefully and experimentally to achieve the best results
- Increasing the input space and the parameters that affects the rainfall prediction.

The primary model was subsequently examined to identify means of enhancing the match between the anticipated and real FR values, through a number of additional steps, for example:

- ➢ Repeat the investigation into the climate of the study area in order to:
  - o Add or remove parameters (input variables);
  - o reformulate the rules based on the climate study (adjusting rules conclusions);
- ➢ Amend the input and/or the output membership functions:
  - o change the centers of the membership functions (change the range, change the type of membership function);
  - o add or remove membership functions;
- ➢ Study the effects of each rule on the model results, selecting those that are of use.

3.6 Final fuzzy models

Following numerous experiments to enhance the primary models, in which over 60 models were tested, a final model was established that carried an acceptable error rate. These four models cover the four regions. The final model included four input variables—wind speed (SW), cloud cover (CL), humidity (HU), and temperature at a specific time (T)—and one output variable, which is estimated rainfall. The input variables were translated to their corresponding membership function (MF) value by establishing parameter types, ranges, and rules to identify all the membership values for any specific input variable. A detailed description of the final model for the middle region is provided in the Figs. (7-11) and Table 6.

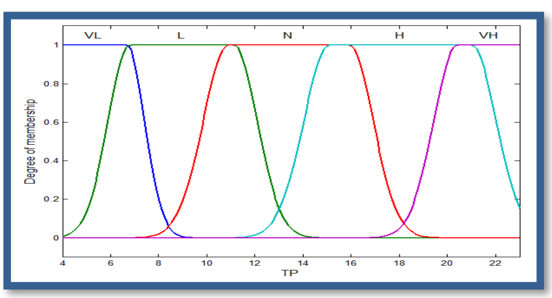


Fig. 7. Membership functions related to temperature (T) – middle model

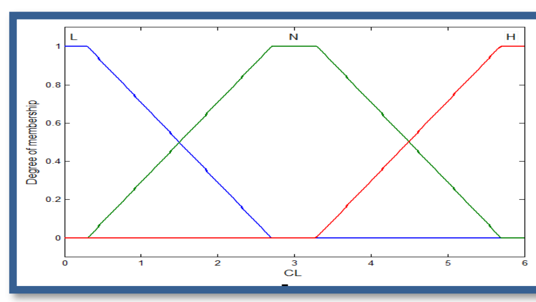
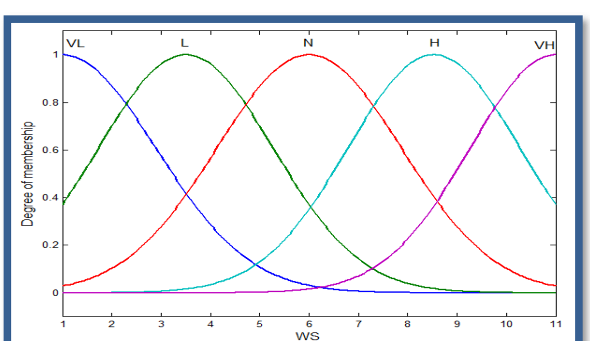


Fig. 8. Membership functions related to cloud (CL) – middle model



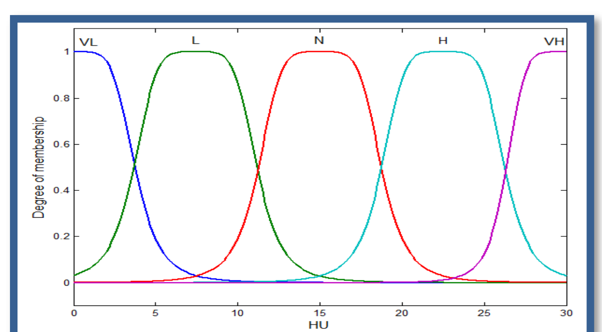


Fig. 9. Membership functions related to wind speed (SW) – middle model

Fig. 10. Membership functions related to humidity (HU) – middle model

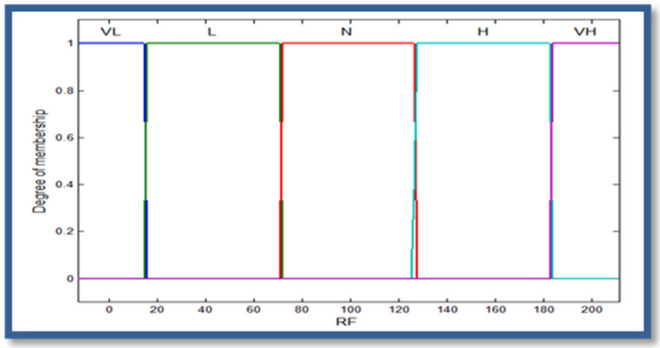


Fig. 11. Membership functions related to rainfall (FR) – middle model

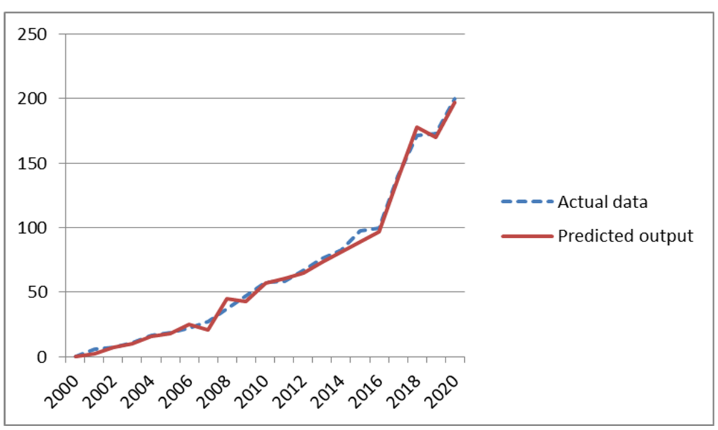
Table 6
Sample of fuzzy rules for the middle region model

1. If (T is VL) and (CL is H) and (SW is H) and (HU is VH) then (FR is VH)
2. If (T is VL) and (CL is H) and (SW is VH) and (HU is H) then (FR is VH)
3. If (T is L) and (CL is N) and (SW is L) and (HU is VL) then (FR is VL)
4. If (T is L) and (CL is N) and (SW is L) and (HU is L) then (FR is VL)
5. If (T is L) and (CL is N) and (SW is L) and (HU is H) then (FR is L)
6. If (T is L) and (CL is N) and (SW is H) and (HU is VL) then (FR is L)
7. If (T is L) and (CL is N) and (SW is VH) and (HU is L) then (FR is L)
8. If (T is L) and (CL is H) and (SW is H) and (HU is L) then (FR is H)
9. If (T is VL) and (CL is N) and (SW is L) and (HU is L) then (FR is L)
10. If (T is L) and (CL is N) and (SW is H) and (HU is L) then (FR is VL)

4. Results and discussion

4.1 The result of the Middle Region model

The results of the Middle Region model and the performance measures are shown are depicted in Fig. 12.



Average PE =**0.**479063543

RMSE = 13.48202

Accuracy (%) = 86.51798

Average MAE = 9.053239

Fig. 13. Actual Values vs. Predicted Values (middle model)

The accuracy of the proposed model has increased significantly after taking into account more climatic parameters that affect rainfall rate, namely: temperature, wind speed, cloud cover, humidity, and carefully constructing the fuzzy sets and rules. A careful reading of the metrics values indicates that the proposed method has successfully modeled the season's rainfall in the middle region of Jordan.

4.2 The result of the North, South and Desert Region model

To summarize, we grouped final results of the proposed model for remaining regions: North, South and Desert regions.

Table 7
Calculated Error Measures for north , south, desert models

Region	Average PE	RMSE	Accuracy (%)	Average MAE
North	0.43	26.56	73.43	16.22
South	0.57	5.85	94.14	4.23
Desert	0.67	3.67	96.32	2.52

A careful reading of the results produces the following conclusions:

- The greater the number of climatic factors affecting rainfall within the model, the greater the accuracy of the results generated. The final model took into account the parameters of temperature, cloud cover, wind speed, and humidity.
- Handling missing values and anomalies found in the dataset is crucial to achieving the best results.
- Using the appropriate order (past values) of the data affects the performance notably.
- The establishment of a separate model with respect to each region with its own climatic parameters is the preferred approach. In this case, the overall territory of study (Jordan) was divided into four regions, and accordingly, four separate fuzzy models were designed.
- Within the fuzzy model, the various parameters for the four proposed regional models were initially established and subsequently adjusted experimentally and heuristically. Consequently, each model differs from all the others in respect of the parameters utilized, including with regard to fuzzy variable values and ranges, membership functions, and fuzzy roles.
- Seasonal rainfall is extremely variable in the middle, the north, and the desert regions.
- The results obtained indicate the models proposed are effective in predicting seasonal rainfall in Jordan.

5. Conclusion

This study sought to forecast seasonal rainfall in Jordan using a fuzzy-based model. The evaluation of the proposed fuzzy model utilized performance metrics including PD, RMSE, MAE, and accuracy. The error measures thus generated suggest that the proposed model is both reliable and acceptable—and is therefore capable of use as a seasonal rainfall prediction tool. The area of study—Jordan—was divided into four regions, each with a similar climate profile. Frequent adjustment of the fuzzy model parameters was undertaken to enhance performance and achieve acceptable results. Following numerous experiments, the climatic parameters of temperature, wind speed, cloud cover, and humidity were adopted as the inputs to the model. The models were flexible and capable of representing a weakly defined link between an input and an output variable. Based on the findings of this study, it is possible to conclude that the fuzzy technique is capable of providing accurate general rainfall predictions. In terms of future work, further optimization may be achieved by combining FL techniques with an alternative method such as artificial neural networks (ANN), exploring the potential of neuro-fuzzy algorithms to provide enhanced results. In addition, experimentation involving larger data sets and more numerous rainfall parameters integrated within the models may prove to be of further value. Finally, it may be useful to explore the utilization of deep-learning techniques within the prediction models in order to manage large data volumes.

References

Agboola, A. H., Gabriel, A. J., Aliyu, E. O., & Alese, B. K. (2013). Development of a fuzzy logic based rainfall prediction model. *International journal of Engineering and Technology*, *3*(4), 427-435.

Al-Ansari, N., Ibrahim, N., Alsaman, M., & Knutsson, S. (2014). Water demand management in Jordan. *Engineering*, *6*(1), 19-26.

Janarthanan, R., Balamurali, R., Annapoorani, A., & Vimala, V. (2021). Prediction of rainfall using fuzzy logic. *Materials Today: Proceedings*, *37*, 959-963.

Jordanian ministry of water & irrigation publication (2015). *Jordan Water Sector: Facts & Figures*.

Kasabov, N. K. (1998). Foundations of neural networks, fuzzy systems, and knowledge engineering. Cambridge, Massachusetts: A Bradford Book.

Kumar Singla, M., Deep Kaur, H., & Nijhawan, P. (2019). Rain Prediction using Fuzzy Logic. *International Journal of Engineering and Advanced Technology, 9*(1).

Logic, F., Novák, V., Perfilieva, I., & Modkof, J. (1999). Mathematicalprinciples of fuzzy logic Dodrecht.

Raddad, K. (2005). Water supply and water use statistics in Jordan. *Proceedings of the IWG-ENV, International Work Session on Water Statistics, Vienna, Austria*, 20-22.

Rahman, M. A. (2020). Improvement of Rainfall Prediction Model by Using Fuzzy Logic. *American Journal of Climate Change*, *9*(04), 391.
Ramadoss, R., Santhanam, K., Shanmugavel, D., & Reddy, K. M. (2020). A study on ductility behaviour of energy efficient mud mortar walls. *Journal of Green Engineering*, *10*(1), 206-224.

Zadeh, L.A. (1965). *fuzzy sets.*
Zahran, B., Mesleh, A., Matouq, M., Alheyasat, O., & Alwada'n, T. (2015). Rainfall prediction in semi-arid regions in Jordan using back propagation neural networks. *International Journal on Engineering Applications*, *3*(6), 158-162.