Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Fuzzy Logic and Machine Learning Approaches for Modeling Monthly Rainfall from Meteorological Variables in Sulaymaniyah, Kurdistan Region of Iraq

Shagul F. Karim
University of Sulaimani
Hadi H. Muhammed
Osnabrück University
Aram S. Naje
University of Sulaimani
Mahtab H. Salih
University of Sulaimani
Mohammed A. Fattah
muhammad.fattah@univsul.edu.iq
University of Sulaimani

Research Article

Keywords: rainfall, fuzzy inference system, machine learning, random forest, relative humidity, skill score

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11421016/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Fuzzy Logic and Machine Learning Approaches for Modeling Monthly Rainfall from Meteorological Variables in Sulaymaniyah, Kurdistan Region of Iraq

Shagul F. Karim[1] Hadi H. Muhammed[2] Aram S. Naje[1] Mahtab H. Salih [1] Mohammed A. Fattah[1]*

[1]College of Agricultural Engineering Sciences, University of Sulaimani, Sulaymaniyah, Kurdistan Region, Iraq.

[2] Institute of Geography, Osnabrück University of Applied Sciences, Osnabrück, Germany.

*Corresponding Author Email: muhammad.fattah@univsul.edu.iq

*Corresponding Author ORCID: https://orcid.org/0000-0002-3726-6182

Abstract

Monthly rainfall information is essential for water management and rainfed agriculture in semi-arid Kurdistan Region of Iraq. This study developed a Mamdani fuzzy inference system (FIS-RH-T-E) in MATLAB that estimates monthly rainfall in Sulaymaniyah from relative humidity (RH), temperature, and evaporation, using 140 expert rules and data from Bakrajo station (2010–2024, n = 180). The FIS was compared with multiple linear regressions (MLR), random forest, gradient boosting, support vector regression, a multilayer perceptron and monthly climatology, evaluated by leave-one-year-out cross-validation and a 2020-2024 hold-out, including one-month-ahead prediction. Climatology alone achieved a Nash-Sutcliffe efficiency (NSE) of 0.45. The FIS reached a coefficient of determination (R^2) of 0.63 and NSE = 0.63 (Kling-Gupta efficiency 0.71), second only to MLR (R^2 = 0.67, NSE = 0.66) and ahead of support vector regression, gradient boosting and random forest (NSE 0.61–0.62); differences among these five models were not significant, whereas all clearly outperformed climatology and multilayer perceptron (NSE = 0.44). Scored on their training data, the machine-learning models reached NSE up to 0.84, which overstates their accuracy. RH dominated both FIS response surfaces and data-driven analyses (anomaly correlation with rainfall 0.63, against −0.21 for temperature and evaporation). All models underestimated the wettest months; the FIS never exceeded 221 mm, whereas observed rainfall reached 316 mm. No model predicted rainfall one month ahead better than climatology. An expert-defined FIS therefore provides transparent estimates as accurate as machine learning for filling gaps in rainfall records, whereas forecasting requires large-scale predictors.

Keywords: rainfall; fuzzy inference system; machine learning; random forest; relative humidity; skill score.

1. Introduction

Rainfall is the main source of water for agriculture in Kurdistan Region of Iraq. In Sulaymaniyah Governorate, winter wheat and barley are largely rainfed, while vegetables and orchards depend on springs and groundwater recharged by winter rain (Najmaddin et al. 2017). Annual rainfall averages 700 mm but varies strongly between years and within the wet season. Dry spells alternate with short, intense storms that exceed soil infiltration capacity and cause runoff, erosion and flash floods. Reliable monthly rainfall information is therefore needed for irrigation scheduling, reservoir operation, crop planning and flood preparedness (Fattah et al. 2020; Hamaamin 2025).

Two practical needs arise from this setting. The first is estimation; many stations in the region have gaps in their rainfall records, while air temperature, relative humidity and pan evaporation are measured more continuously. A model that estimates monthly rainfall from these routinely observed variables can fill such gaps and extend records. The second is prediction; water managers need rainfall information before the month begins, which requires inputs available at the time of issue. The two tasks are often conflated in the literature, although a model that performs well when fed concurrent observations may have little skill with a one-month lead.

Statistical, stochastic and machine-learning (ML) models have been widely used for both tasks (Sham et al. 2025). In Sulaymaniyah, recent studies applied data-mining, artificial neural network and deep-learning methods to daily precipitation and climate variables (Hamaamin 2025; Hamaamin et al. 2025). Even well-designed models cannot approach perfect accuracy, because atmospheric processes are dynamic and highly variable (Małolepsza et al. 2025). ML models can be accurate, but their internal logic is difficult to inspect. For decisions that affect water allocation and farmer livelihoods, transparent models whose reasoning can be checked against physical understanding are preferable when they perform comparably (Rudin 2019).

Fuzzy inference systems (FIS) offer such transparency. Based on fuzzy set theory (Zadeh 1965), they represent the imprecision of environmental measurements, which differs from random variability (Jantakoon 2016). A Mamdani FIS maps inputs to outputs through linguistic IF–THEN rules such as IF humidity is high AND evaporation is low THEN rainfall is high (Ross 2009). Mamdani systems have been applied widely in hydrology and engineering (Katambara and Ndiritu 2009; Tosun et al. 2011; Akgun et al. 2012), including the assessment of treated wastewater for irrigation in Sulaymaniyah (Hamabor et al. 2023). For rainfall, they have been used in the USA (Hasan et al.

2008), Iran (Fallah-Ghalhary et al. 2009), Egypt (Asklany et al. 2011), Malaysia (Mohd Safar et al. 2019), India (Janarthanan et al. 2021; Selvaraj et al. 2023) and Indonesia (Damayanti et al. 2022), among others (Rahman 2020). Most of these studies share three limitations. First, membership functions and rules are set by expert judgment and evaluated on the same data used to design them, thus reported accuracy is in-sample. Second, performance is rarely compared with a climatological baseline or with ML models. In climates with a strong annual cycle, even the long-term monthly mean explains much of the rainfall variance, hence a high coefficient of determination does not demonstrate skill. Third, the effect of centroid defuzzification on extreme values is seldom examined, although it restricts predictions to the interior of the output universe and systematically underestimates the largest events.

To our knowledge, no previous study has directly compared a rule-based fuzzy rainfall model for Sulaymaniyah with machine-learning models under a common, independent evaluation. This study addresses that gap. The objectives were to: (i) develop a Mamdani FIS in MATLAB that estimates monthly rainfall from relative humidity (RH), air temperature (T) and pan evaporation (E); (ii) interpret the FIS through its rule viewer and response surfaces; (iii) compare its performance with multiple linear regression, random forest, gradient boosting, support vector regression, a multilayer perceptron and monthly climatology, using leave-one-year-out and chronological validation; and (iv) test whether the same variables can predict rainfall one month ahead. We hypothesized that the FIS would perform comparably to the machine-learning models while offering readable rules, and that skill would largely vanish when concurrent inputs are replaced by lagged ones.

2. Materials and Methods

2.1 Study area

Sulaymaniyah lies on a high plateau at about 850–900 m a.s.l. in the north-east of Iraq, surrounded by Goyzha, Azmir, Glazarda and Piramagrun ranges (Figure 1). The governorate extends between 34°32′–36°35′ N and 44°30′–46°15′ E and covers about 18500 km^2. The climate is semi-arid, Irano-Turanian, with cold, wet winters and long, hot, dry summers (Hamaamin et al. 2025). Winter rainfall is delivered mainly by eastward-moving Mediterranean cyclones, with additional moisture from the Black Sea, Red Sea and Persian Gulf (Mohammadzadeh et al. 2020; Al-Naseri et al. 2022; Abdulrahman 2023).

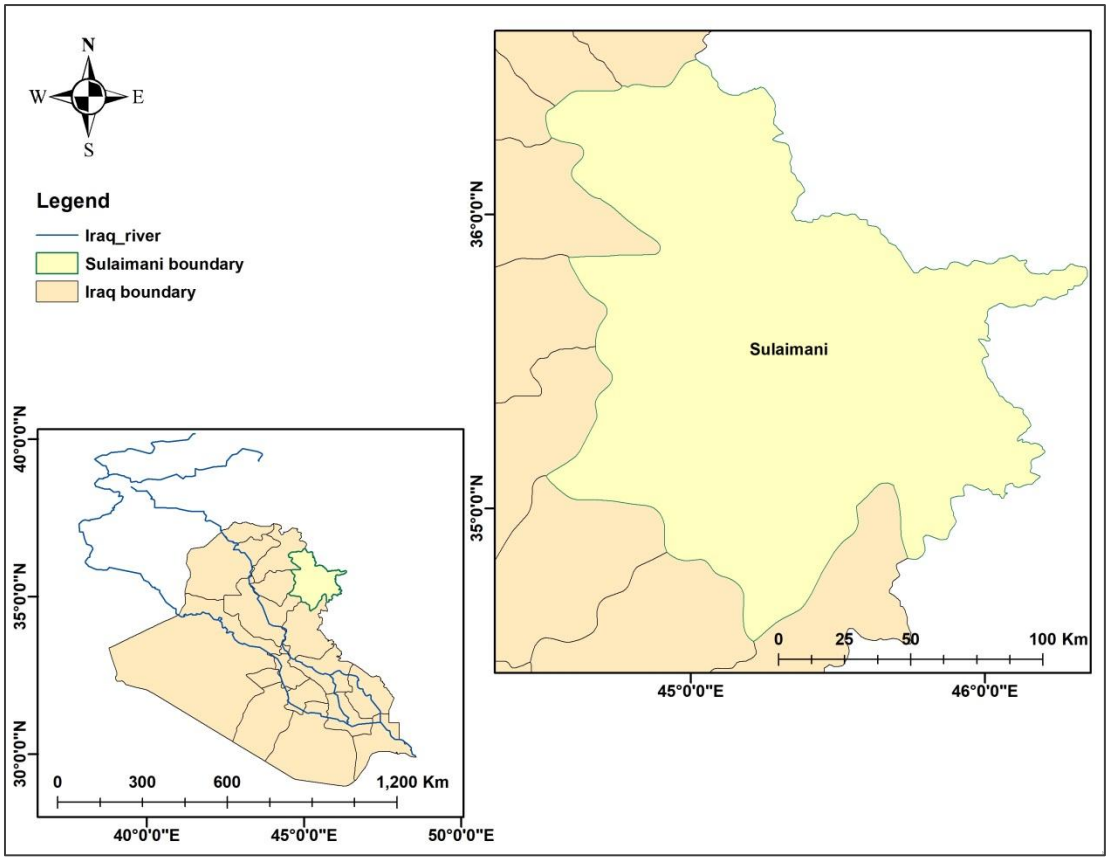


Fig. 1 Location of Sulaymaniyah Governorate in the Kurdistan Region of Iraq.

2.2 Data Collection and Quality Control

Monthly data for January 2010 to December 2024 (n = 180) were obtained from Bakrajo agrometeorological station (35°33′ N and 45°21′ E, 850–900 m a.s.l.), operated by Iraqi Agrometeorological Network. The variables were monthly rainfall total P (mm month^{-1}), and monthly means of daily air temperature T (°C), relative humidity RH (%) and Class-A pan evaporation E (mm d^{-1}). Snowfall was recorded as water equivalent. The series were checked for missing and duplicated months, physically implausible values, and runs of identical values that might indicate gap-filling. No months were missing. Fifty months had no rainfall, 48 of them between June and September. Mean temperatures of June–August 2012 (36.9–40.6 °C) exceeded the 2010–2024 means of those months by more than three standard deviations and were checked against the station records. The core wet season was defined as November–April, which contributes 89.8% of annual rainfall; May–October were treated as dry and transitional season.

Pearson and Spearman correlations between each input and rainfall were computed, together with partial correlations controlling for the other two inputs. Because all four variables follow the same annual cycle, correlations were also computed on monthly anomalies (value minus the long-term mean for that calendar month), for all months and for the wet season. Multicollinearity among the input variables was assessed using the variance inflation factor (VIF).

2.3 Fuzzy Inference System (FIS)

A fuzzy inference system (FIS), based on fuzzy set theory introduced by Zadeh (1965), was used to represent relationships among variables characterized by uncertainty, vagueness, and nonlinearity. Within the FIS framework, numerical (crisp) input variables were transformed into fuzzy sets using membership functions, which assigned each input a membership degree ranging from 0 to 1 (Katambara and Ndiritu 2009: Mohd Safar et al. 2019). The membership function is expressed as:

$$A = \{x, \mu_A(x) | x \in U\}, \qquad \mu_A(x) \in [0,1] \dots\dots\dots\dots (1)$$

Where $\mu_A(x)$ represents the membership degree of element x in fuzzy set A, and U denotes the universe of discourse. Membership functions can take different forms, including triangular, trapezoidal, and Gaussian shapes. The FIS process consists of four main stages: fuzzification, rule evaluation, aggregation, and defuzzification (Figure 2). During fuzzification, crisp input values are converted into their corresponding membership degrees. The inference engine then evaluates the predefined IF-THEN rules based on the fuzzified inputs. The resulting fuzzy outputs from the activated rules are aggregated into a single fuzzy output and subsequently converted into a crisp numerical value through defuzzification. In the present study, a Mamdani-type FIS was adopted (Ross 2009; Hamabor et al. 2023).

$$R_i: IF\ x_1\ is\ A_{i1}\ AND\ x_2\ is\ A_{i2}\ AND\ x_3\ is\ A_{i3}\ THEN\ y\ is\ B_i \dots\dots\dots\dots (2)$$

Where R_i represents the i^{th} fuzzy rule, x_1, x_2, and x_3 represent the input variables, A_{i1}, A_{i2}, and A_{i3} represent their corresponding fuzzy sets, and B_i denotes the fuzzy output set. Each rule R_i defines a specific input–output relationship within the FIS (Tosun et al. 2011; Akgun et al. 2012).

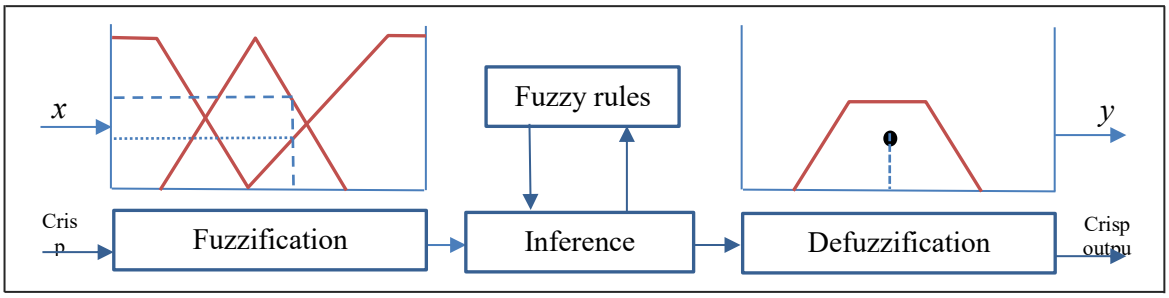


Fig. 2 Steps of fuzzy inference system (FIS).

In this study, the Fuzzy Logic Toolbox in MATLAB was used to develop the FIS for rainfall prediction. The structure of the proposed model is presented in Figure 3. The model, designated FIS-RH-T-E, was developed using three meteorological variables as input variables: relative humidity (RH), temperature (T), and evaporation (E). These variables were selected to represent important atmospheric conditions associated with rainfall occurrence and intensity. In the first step, the observed values of RH, T, and E were simultaneously introduced into the FIS as input variables. The numerical values of these variables were then converted into fuzzy linguistic representations through predefined membership functions.

During fuzzification, the numerical values of relative humidity, temperature, and evaporation were assigned degrees of membership to their respective fuzzy sets. The resulting fuzzy inputs were then evaluated using a rule base established from the relationships among the three meteorological variables and rainfall. The outputs of the activated rules were aggregated to form the final fuzzy output, which was subsequently converted into a crisp rainfall value through defuzzification. For example, as illustrated in Figure 3, input values of 57.2% relative humidity, 21.5°C temperature, and 4.1 mm d^{-1} evaporation produced a predicted rainfall value of 64.48 mm.

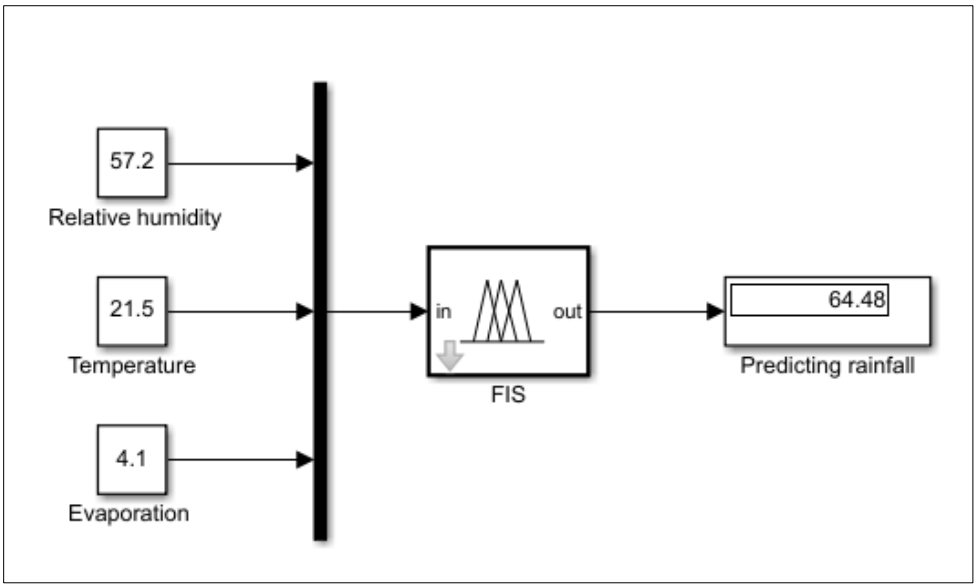


Fig. 3 Structure of the MATLAB-based fuzzy inference system (FIS-RH-T-E) for rainfall prediction.

The FIS-RH-T-E was constructed using different numbers of membership functions for each input variable. Relative humidity was represented by seven membership functions: very low (VLRH), low (LRH), normal (NRH), high (HRH), very high (VHRH), very very high (VVHRH), and extremely high (VVVHRH). Temperature was represented by five membership functions: very low (VLT), low (LT), normal (NT), high (HT), and very high (VHT). Evaporation was represented by four membership functions: low (LE), normal (NE), high (HE), and very high (VHE). The ranges of the input variables were defined based on the observed meteorological data and expert knowledge, following the fuzzy rainfall modeling approaches of Hasan et al. (2008), Sojitra et al. (2015) and Selvaraj et al. (2023). The membership functions and their corresponding fuzzy sets are presented in Figure 4.

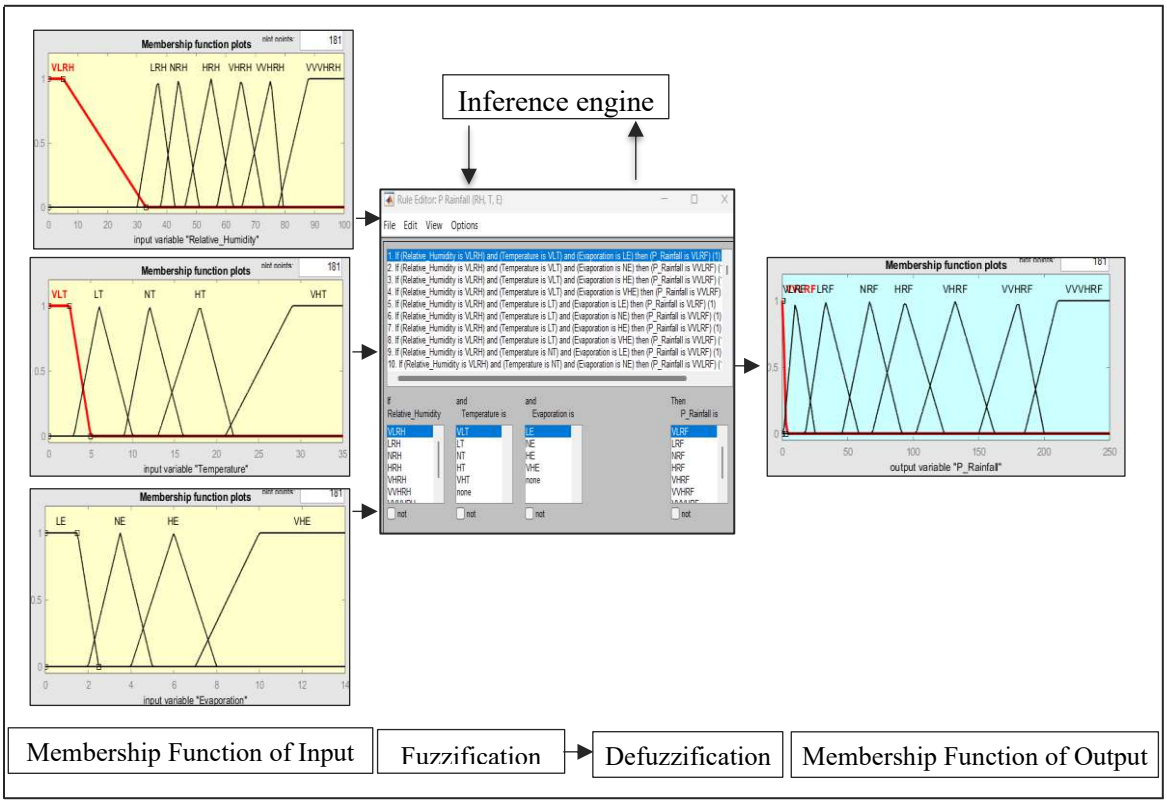


Fig. 4 Membership functions and fuzzy sets of the input and output variables used in the FIS-RH-T-E rainfall prediction model.

The rainfall output variable was represented by eight membership functions (fuzzy sets): very very low rainfall (VVLRF), very low rainfall (VLRF), low rainfall (LRF), normal rainfall (NRF), high rainfall (HRF), very high rainfall (VHRF), very very high rainfall (VVHRF), and extremely high rainfall (VVVHRF). The total number of possible fuzzy inference rules was determined from the number of membership functions assigned to the three input variables. With seven membership functions for relative humidity, five for temperature, and four for evaporation, the total number of possible rule combinations was calculated as: $R = 7 \times 5 \times 4 = 140$ rules

Accordingly, the FIS-RH-T-E consisted of 140 fuzzy inference rules, representing combinations of relative humidity, temperature, and evaporation conditions and their corresponding rainfall output fuzzy sets. The fuzzy rules were developed based on the researchers' knowledge and expertise in rainfall prediction, supported by the approach described by Asklany et al. (2011). The resulting rule base was used to infer the rainfall output from different

combinations of the three meteorological input variables. The complete structure of the FIS and the membership functions of the input and output variables are presented in Figs. 4 and 5, respectively.

For defuzzification, the centroid (center of gravity) method was used to convert the aggregated fuzzy output into a single crisp value representing the predicted rainfall. The method is expressed as follows:

$$Z_{centroid} = \frac{\int_z \mu_A(z) z dz}{\int_z \mu_A(z) dz} \dots\dots\dots\dots (3)$$

Where $Z_{centroid}$ is the resulting crisp rainfall value and $\mu_A(z)$ is the aggregated output membership function.

2.4 Machine learning and reference models

To place the FIS-RH-T-E in context, five data-driven models were fitted to the same three inputs in Python. Each model estimates the monthly rainfall $\hat{P}$ from the input vector x = (RH, T, E). For SVR and the MLP, the inputs were first standardized:

$$z_j = \frac{x_j - \bar{x}_j}{s_j} \dots\dots\dots\dots (4)$$

where $\bar{x}_j$ and s_j are the mean and standard deviation of input j in the training years. Negative predictions were set to zero. Hyperparameters were selected by five-fold cross-validation grouped by year, nested inside each training set, therefore the test year was never used for model fitting or tuning.

Multiple linear regression (MLR) assumes a linear relationship between rainfall and the inputs.

$$\hat{P} = b_0 + b_1 RH + b_2 T + b_3 E \dots\dots\dots\dots (5)$$

whose coefficients b_0–b_3 minimize the sum of squared errors over the n training months:

$$\min_{b_0,\dots,b_3} \sum_{i=1}^{n} \left(P_i - \hat{P}_i\right)^2 \dots\dots\dots\dots (6)$$

Random forest (RF) averages the predictions of B regression trees (Breiman 2001)

$$\hat{P}(\mathrm{x}) = \frac{1}{B} \sum_{b=1}^{B} T_b(\mathrm{x}) \dots\dots\dots\dots (7)$$

where each tree T_b divides the input space by successive binary splits (e.g., RH > 55%) chosen to minimise the squared error, and predicts the mean training rainfall within each resulting region. Each tree is grown on a bootstrap

sample of the training months and considers a random subset of the inputs at each split. B = 500; the minimum number of months per leaf (1, 3 or 5) and the number of inputs considered per split (1–3) were tuned.

According to Friedman (2001), gradient boosting regression (GBR) builds an additive model sequentially by fitting small regression trees to the residuals of the current model, thereby progressively reducing prediction errors.

$$F_m(\mathrm{x}) = F_{m-1}(\mathrm{x}) + \nu\, h_m(\mathrm{x}), \quad m = 1, \ldots, M \ldots\ldots\ldots\ldots (8)$$

starting from the mean training rainfall F_0. Here h_m is a regression tree fitted to the residuals $P_i - F_{m-1}(x_i)$, $\nu = 0.03$ is the learning rate and M = 300; each tree was fitted on a random 80% of the training months, its depth (1–3) was tuned, and the prediction is $\hat{P} = F_M(\mathrm{x})$.

As proposed by Smola and Schölkopf (2004), support vector regression (SVR) estimates rainfall as a weighted sum of similarities between the standardised input vector **z** of the target month and those of the training months, z_i, while controlling model complexity to minimize prediction errors.

$$\hat{P}(\mathrm{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\, K(\mathrm{z}_i, \mathrm{z}) + b \ldots\ldots\ldots\ldots (9)$$

using the radial basis function kernel

$$K(\mathrm{z}_i, \mathrm{z}) = \exp\left(-\gamma \|\mathrm{z}_i - \mathrm{z}\|^2\right) \qquad (10)$$

The weights α_i, α_i^* and the intercept b result from solving

$$\min \frac{1}{2}\|\mathrm{w}\|^2 + C \sum_{i=1}^{n} \max\left(0,\ \left|P_i - \hat{P}_i\right| - \varepsilon\right) \ldots\ldots\ldots\ldots (11)$$

where w is the weight vector in the kernel feature space, errors smaller than ε are ignored, and C controls the trade-off between fitting the data and keeping the function smooth. C (10, 100 or 1000) and ε (1, 5 or 10 mm) were tuned, and γ = 1/3 (the inverse of the number of standardised inputs).

The multilayer perceptron (MLP) is a feed-forward neural network with two hidden layers of 16 and 8 neurons:

$$\mathrm{h}_1 = \mathrm{ReLU}(\mathrm{W}_1\mathrm{z} + \mathrm{c}_1), \quad \mathrm{h}_2 = \mathrm{ReLU}(\mathrm{W}_2\mathrm{h}_1 + \mathrm{c}_2), \quad \tilde{P} = \mathrm{w}_3^{\mathrm{T}}\mathrm{h}_2 + c_3 \ldots\ldots\ldots\ldots (12)$$

where $\mathrm{ReLU}(u) = \max(0, u)$, W_k and c_k are weight matrices and bias vectors (209 parameters in total), and $\tilde{P}$ is the standardized rainfall, converted back to mm after prediction. The weights minimize the penalized squared error

$$L = \frac{1}{2n}\sum_{i=1}^{n}\left(\tilde{P}_i - \hat{\tilde{P}}_i\right)^2 + \frac{\lambda}{2n}\sum_{k}\|\mathrm{W}_k\|^2 \ldots\ldots\ldots\ldots (13)$$

with λ = 0.01, using the L-BFGS algorithm (at most 2000 iterations). The reported MLP prediction is the mean of five networks trained from different random initial weights.

Finally, monthly climatology served as a baseline that represents the skill obtainable from the seasonal cycle alone. For calendar month m it predicts the mean observed rainfall of that month over the set Y of N_Y training years:

$$C_m = \frac{1}{N_Y} \sum_{y \in Y} P_{y,m} \quad \ldots\ldots\ldots\ldots (14)$$

The FIS-RH-T-E was not refitted for the comparison; its monthly predictions were computed with evalfis in MATLAB and scored on exactly the same months as the other models.

2.5 Evaluation design

The data-driven models and climatology were evaluated by leave-one-year-out (LOYO) cross-validation, a form of cross-validation widely used to assess statistical climate prediction models (Michaelsen 1987; Elsner and Schmertmann 1994): each of the 15 years was withheld in turn, the models were fitted on the other 14 years, and the predictions for the withheld years were pooled. Whole years rather than single months were withheld, because months of the same season are not independent and withholding single months would overestimate accuracy (Roberts et al. 2017). A chronological hold-out (for training 2010–2019, for testing 2020–2024) checked performance in the most recent, partly dry years. The FIS-RH-T-E has no parameters fitted to the data, thus its predictions are identical under every scheme and were scored on exactly the same months. However, its membership-function ranges and rules were defined with knowledge of the whole 2010–2024 record. The data-driven models were therefore also fitted and scored on all 180 months (in-sample), which provides the like-for-like reference for the FIS; LOYO results represent accuracy expected for new years. The difference in root-mean-square error (RMSE; Eq. 18) between FIS and each data-driven model was tested with a paired year-block bootstrap (2000 resamples of whole years).

Finally, a one-month-ahead test examined whether the variables can predict next month's rainfall: the data-driven models used RH, T, E and rainfall of month $t - 1$ together with sine and cosine of the targeted month, and were compared with climatology and persistence (rainfall of month t equals that of month $t - 1$). Because the rules of FIS relate concurrent variables, this test was run for the data-driven models only.

2.6 Performance measures

Agreement between observed (P_i) and predicted ($\hat{P}_i$) monthly rainfall over n months was quantified with the coefficient of determination (R^2), the Nash–Sutcliffe efficiency (NSE), the Kling–Gupta efficiency (KGE) as proposed by Gupta et al. (2009), the root-mean-square error (RMSE), the mean absolute error (MAE) and the percent bias (PBIAS), interpreted with the guidance of Moriasi et al. (2007):

$$R^2 = r^2 = \left[\frac{\sum_{i=1}^{n}(P_i - \bar{P})\left(\hat{P}_i - \bar{\hat{P}}\right)}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}\left(\hat{P}_i - \bar{\hat{P}}\right)^2}}\right]^2 \qquad (15)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(P_i - \hat{P}_i\right)^2}{\sum_{i=1}^{n}(P_i - \bar{P})^2} \qquad (16)$$

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{\hat{P}}}{\sigma_P} - 1\right)^2 + \left(\frac{\bar{\hat{P}}}{\bar{P}} - 1\right)^2} \qquad (17)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{P}_i - P_i\right)^2} \qquad (18)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{P}_i - P_i\right| \qquad (19)$$

$$\mathrm{PBIAS} = 100 \times \frac{\sum_{i=1}^{n}\left(\hat{P}_i - P_i\right)}{\sum_{i=1}^{n} P_i} \qquad (20)$$

where $\bar{P}$ and the barred $\hat{P}$ denote mean observed and mean predicted rainfall, r is the Pearson correlation coefficient between observed and predicted rainfall, and σ_P and $\sigma_{\hat{P}}$ are their standard deviations. The Pearson coefficient r is also used in Section 3.1 to describe relationships between variables. R^2 measures only how closely the predictions follow the observations, whereas NSE and KGE also penalise systematic over- or underestimation; NSE = 1 and KGE = 1 indicate perfect agreement, and NSE = 0 means that the model is no better than the mean observed rainfall. PBIAS > 0 indicates overestimation. According to Murphy (1988), skill relative to climatology was assessed using the mean squared error skill score (MSESS) and the anomaly correlation coefficient (ACC):

$$\mathrm{MSESS} = 1 - \frac{\sum_t \left(\hat{P}_t - P_t\right)^2}{\sum_t \left(C_{m(t)} - P_t\right)^2} \ldots\ldots\ldots\ldots (21)$$

$$\mathrm{ACC} = \mathrm{corr}\left(P_t - C_{m(t)},\ \hat{P}_t - C_{m(t)}\right) \ldots\ldots\ldots\ldots (22)$$

where $C_{m(t)}$ is the out-of-sample climatological mean for the calendar month of *t*. MSESS > 0 means a model beats climatology. Following Piñeiro et al. (2008), observed values were regressed on predicted values. Metrics were computed for all months, for the wet season, and for the upper decile of observed rainfall. Ninety-five percent confidence intervals (CI) of the LOYO metrics were obtained with a year-block bootstrap (2000 resamples of whole years), which also gave the probability that MSESS ≤ 0.

2.7 Software

The FIS-RH-T-E was built and evaluated with the Fuzzy Logic Toolbox of MATLAB R2007b (version 2.2.6; The MathWorks Inc. 2007). The data-driven models, the statistical analyses and the figures were produced with Python 3.11.9 (Van Rossum and Drake 2009), using NumPy 2.4.6 (Harris et al. 2020), pandas 3.0.6 (McKinney 2010), SciPy (Virtanen et al. 2020), scikit-learn 1.9.1 (Pedregosa et al. 2011) and Matplotlib (Hunter 2007).

3. Results and Discussion

3.1 Rainfall climatology and relationships among variables

Rainfall showed a strong annual cycle: November–April delivered 89.8% of the mean annual total of 688 mm, while June–September received less than 3 mm month^{-1} on average (Figure 5; Table, 1). Monthly rainfall ranged from 0 to 316.1 mm (December 2018), with a coefficient of variation of 62.7% in wet season, and annual totals ranged from 292 mm (2021) to 1051 mm (2019). T and E peaked in July–August (34.5 °C and 12.3 mm d^{-1}), and mean RH ranged from 24.8% in July to 71.1% in January.

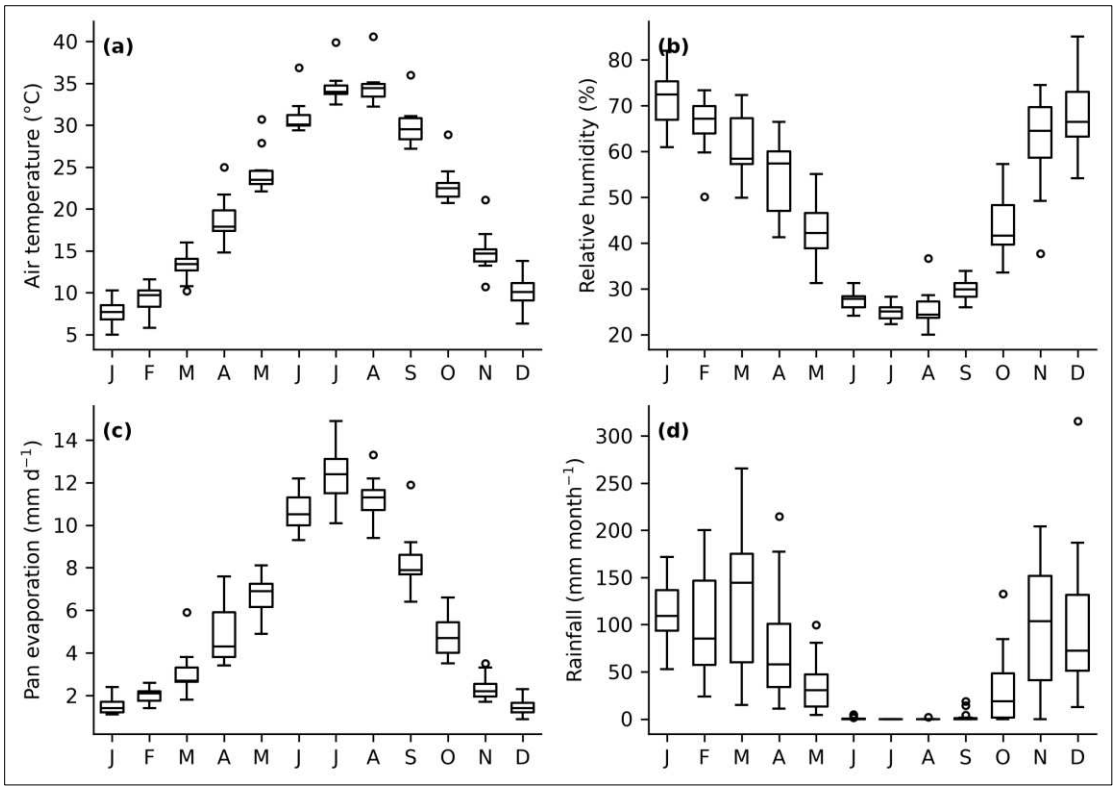


Fig. 5 Monthly distribution of (a) air temperature, (b) relative humidity, (c) pan evaporation and (d) rainfall at the Bakrajo station, 2010–2024. Boxes show the interquartile range, lines the median.

[Insert Table 1 here]

Table 1 Descriptive statistics of the monthly station data, 2010–2024.

Subset	Variable	Unit	Mean	SD	CV (%)	Min	Max
All	T	°C	20.9	9.7	46.6	5.0	40.6
All	RH	%	48.1	18.3	37.9	20.0	85.1
All	E	mm d^{-1}	5.8	3.9	68.3	0.9	14.9
All	P	mm $month^{-1}$	57.3	66.9	116.7	0.0	316.1
Nov–Apr	T	°C	12.3	4.2	34.0	5.0	25.0
Nov–Apr	RH	%	63.9	9.3	14.6	37.7	85.1
Nov–Apr	E	mm d^{-1}	2.5	1.4	55.3	0.9	7.6
Nov–Apr	P	mm $month^{-1}$	103.0	64.6	62.7	0.0	316.1

SD: standard deviation; CV: coefficient of variation.

All three inputs were strongly correlated with rainfall over all months (r = 0.80 for RH, −0.70 for T and −0.68 for E (Table, 2). They were also very strongly correlated with each other: RH was negatively correlated with T (r = −0.94) and E (r = −0.92), and T and E positively with each other (r = 0.94), indicating that all three inputs largely carry the same seasonal information, with VIF values ranging from 9.2 to 12.2 (Table, 2). After removing the mean seasonal cycle, correlations with rainfall anomalies fell to r = 0.63 (RH), −0.21 (T) and −0.21 (E) over all months, and to 0.64, −0.29 and −0.24 in wet season. Controlling for other two inputs, only RH retained a significant partial correlation with rainfall (r = 0.59, $p < 0.001$; T: 0.13, p = 0.08; E: 0.14, p = 0.06). Most of the association of T and E with rainfall therefore reflects the shared annual cycle, whereas RH carries additional month-to-month information.

[Insert Table 2 here]

Table 2 Correlation of each input with monthly rainfall (columns 2–6) and with the other two inputs (columns 7–8). Partial r controls for the other two inputs; anomalies are departures from the calendar-month mean; p values in brackets; VIF: variance inflation factor.

Input	Pearson r	Spearman ρ	Partial r	Anomaly r (all)	Anomaly r (wet)	r with other inputs	VIF
RH	0.80	0.89	0.59 (<0.001)	0.63 (<0.001)	0.64 (<0.001)	T: −0.94; E: −0.92	9.4
T	−0.70	−0.82	0.13 (0.08)	−0.21 (0.004)	−0.29 (0.006)	RH: −0.94; E: 0.94	12.2
E	−0.68	−0.82	0.14 (0.06)	−0.21 (0.004)	−0.24 (0.02)	RH: −0.92; T: 0.94	9.2

3.2 FIS Rule Viewer Analysis

The fuzzy rule-based rainfall prediction system (FIS-RH-T-E) produced rainfall predictions by considering simultaneous conditions of relative humidity, temperature, and evaporation. Figure 6 illustrates the response of the system for relative humidity of 57.2%**,** temperature of 21.5°C, and evaporation of 4.1 mm d^{-1}. Under these input conditions, predicted rainfall was 64.48 mm, which was displayed as approximately 64.5 mm in Rule Viewer. The figure shows that the three meteorological variables contributed simultaneously to rainfall output through activated fuzzy rules. Thus, the predicted rainfall was determined from combined atmospheric conditions rather than from the effect of an individual variable alone.

This result is in agreement with Rahman (2020), who demonstrated that fuzzy logic can be used to predict rainfall by combining meteorological inputs through fuzzy membership functions and IF–THEN rules. A similar expert-rule-based approach was reported by Janarthanan et al. (2021), who used temperature and wind speed as fuzzy inputs and converted the resulting fuzzy inference into quantitative rainfall predictions. These studies support the use of multiple meteorological variables within a fuzzy rule-based framework for rainfall prediction.

The Rule Viewer also demonstrates the ability of FIS-RH-T-E to handle different degrees of membership among input variables. For the specified values of RH, temperature, and evaporation, several rules were activated with different strengths, resulting in the final rainfall prediction of 64.48 mm. This indicates that rainfall prediction in the developed model depends on the combined state of the three meteorological variables. The result presented in Figure 6 therefore provides an example of how fuzzy rule base converts a particular combination of atmospheric conditions into a quantitative rainfall prediction.

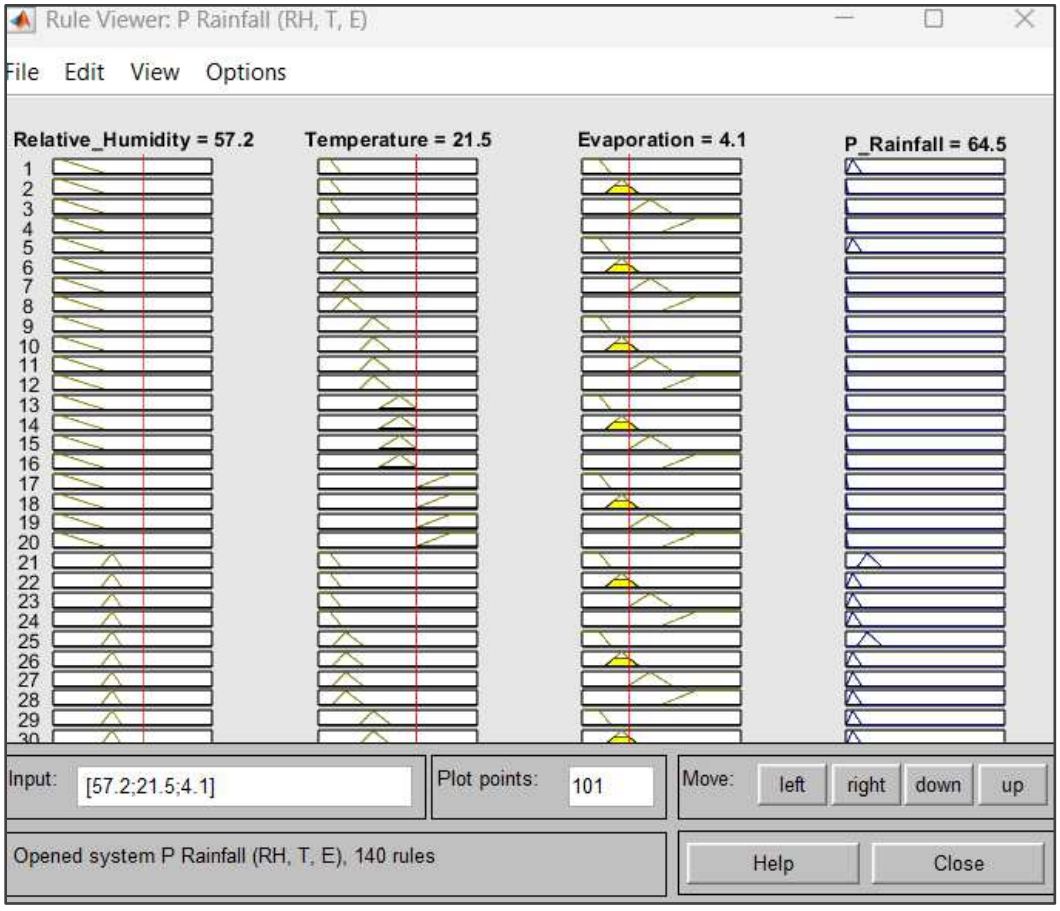


Fig. 6 Fuzzy rule viewer of the FIS-RH-T-E rainfall prediction system.

3.3 Surface Plot Analysis

The relationships between input variables and predicted rainfall were further evaluated using the three-dimensional surface plots presented in Figure 7. The relative humidity–temperature **(**RH–T**)** surface indicates that predicted rainfall was strongly affected by relative humidity. At relatively low RH values, predicted rainfall remained comparatively low over the temperature range, whereas increasing RH resulted in a marked increase in predicted rainfall. The maximum predicted rainfall occurred under combinations involving relatively high RH and suitable temperature conditions, reaching approximately 200–220 mm, while minimum rainfall predictions occurred under conditions associated with low RH. The surface also indicates that the effect of temperature was not constant across the RH range, showing that rainfall response to temperature depended on the prevailing relative humidity conditions.

Relative humidity–evaporation (RH–E) surface showed a pattern similar to that observed for RH–temperature relationship. Predicted rainfall increased considerably with increasing relative humidity, while increasing evaporation was generally associated with lower predicted rainfall under the modeled conditions. The maximum predicted rainfall reached approximately 200–220 mm under conditions of high relative humidity and relatively low evaporation, whereas the lowest rainfall predictions occurred under lower RH conditions. This pattern indicates that high evaporation does not necessarily result in high predicted rainfall. Rather, its effect depends on the atmospheric moisture condition represented by relative humidity. Therefore, the surface demonstrates an important interaction between moisture availability and evaporative demand in the rainfall prediction model.

The use of three-dimensional surface plots to examine interactions among meteorological variables is also consistent with Selvaraj et al. (2023), who applied a Mamdani fuzzy rainfall model and used surface plots to evaluate the variation of rainfall-related outputs with changing combinations of input variables. Although their model structure and input variables differed from the present FIS-RH-T-E, their approach provides methodological support for interpreting fuzzy-model responses through surface analysis.

In contrast, the evaporation–temperature (E–T) surface showed a considerably narrower range of predicted rainfall than the two surfaces involving relative humidity. Predicted rainfall on this surface was generally concentrated between approximately 40 and 60 mm, with the maximum predicted rainfall being close to 60 mm. This difference occurs because the E–T surface represents the response of the three-input FIS while relative humidity is held at a fixed value. In other words, the surface represents $R=F(E,T \mid RH=RH_f)$, whereas the RH–T and RH–E surfaces allow relative humidity to vary. Consequently, E–T surface does not represent the complete range of rainfall outputs

possible from the FIS; it represents rainfall response to evaporation and temperature under one particular humidity condition. The relatively narrow range and maximum value of approximately 60 mm should therefore be interpreted as a conditional response of FIS rather than as a limitation indicating that temperature and evaporation cannot be associated with higher rainfall under other humidity conditions.

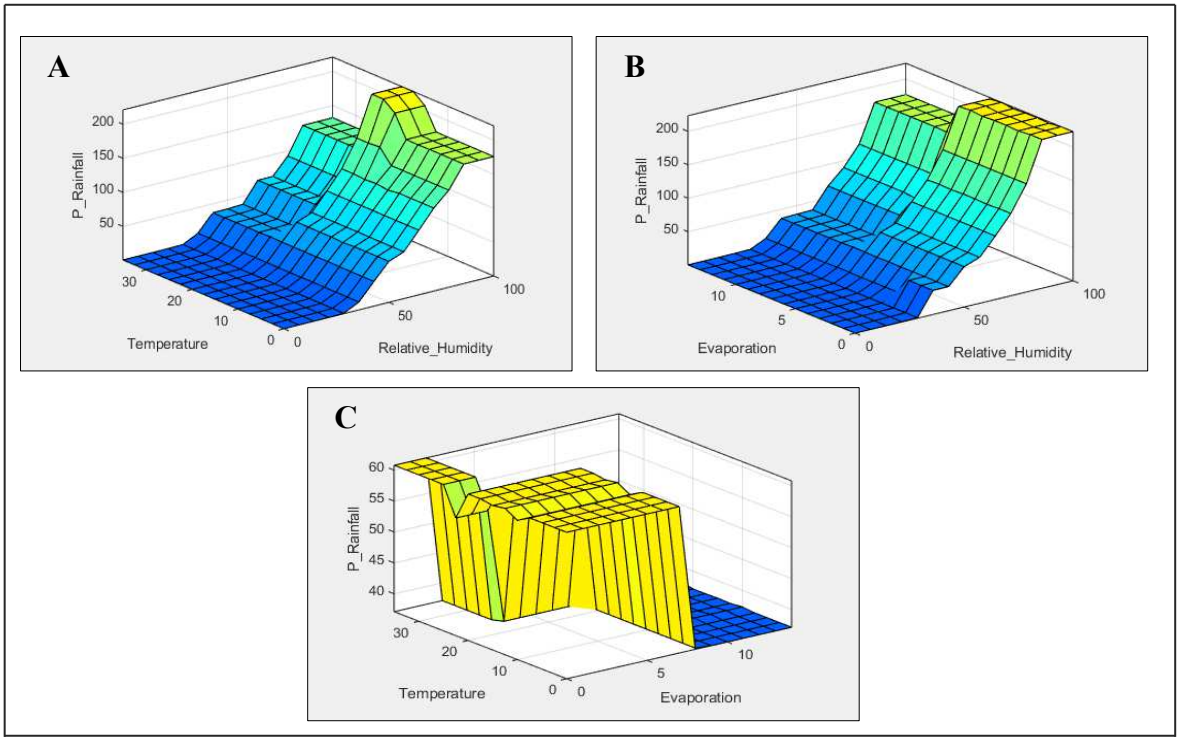


Fig. 7 Three-dimensional surface plots of predicted rainfall using the FIS-RH-T-E model: (a) relative humidity and temperature, (b) relative humidity and evaporation, and (c) evaporation and temperature.

3.4 Interactions among Temperature, Evaporation, and Relative Humidity

The three surface plots collectively indicate that effects of temperature and evaporation on predicted rainfall depended on the prevailing relative humidity conditions. The substantially larger variation observed in the surfaces involving relative humidity indicates that FIS response was strongly dependent on atmospheric moisture conditions and their interaction with temperature and evaporation.

3.5 Relative Humidity as an Important Component of Rainfall Prediction

Among the three input variables, relative humidity showed a particularly strong influence on the predicted rainfall response in the surface plots. Both RH–T and RH–E surfaces showed a substantial increase in predicted rainfall as relative humidity increased. This behavior is physically reasonable because relative humidity represents the degree of atmospheric moisture relative to moisture-holding capacity of the air. Conditions of higher relative humidity indicate a greater degree of atmospheric moisture availability, which can provide more favorable conditions for rainfall formation when other atmospheric conditions are suitable.

The data-driven analyses support this model response. After removing the mean seasonal cycle, RH remained correlated with rainfall anomalies ($r = 0.63$), whereas T and E were only weakly correlated ($r = -0.21$ each), and only RH had a significant partial correlation with rainfall when the other two inputs were held constant $r = 0.59$, $p < 0.001$ (Table, 2). In the random forest, randomly permuting RH increased the out-of-sample RMSE by 51 mm, whereas permuting T or E changed it by about 1 mm.

The relatively strong role of relative humidity in rainfall prediction may be related to regional climatological setting of northern Iraq, where precipitation is strongly influenced by moisture-bearing weather systems originating from the Mediterranean region. Moisture transported from the Mediterranean Sea, together with contributions from other regional moisture sources such as the Black Sea and Persian Gulf, contributes to the atmospheric moisture conditions associated with rainfall over the region.

The strong role of relative humidity in the present model is consistent with previous rainfall studies that have included humidity as an important meteorological input. Rahman (2020), for example, identified relative humidity together with other atmospheric parameters as relevant to daily precipitation variability. Regional evidence also provides a physical context for the present result: Al-Naseri et al. (2022) identified the Mediterranean Sea and Black Sea as major moisture-source regions for Iraqi precipitation, while Mohammadzadeh et al. (2020) reported seasonal contributions from the Mediterranean Sea, Black Sea, Persian Gulf, and Red Sea in precipitation over the west of Iran and east of Iraq. These findings support the interpretation that atmospheric moisture conditions represented by relative humidity are relevant to rainfall formation in the region. However, Al-Naseri et al. (2022) reported a weak negative correlation between relative humidity and precipitation amount at the national scale; therefore, the positive RH response observed here should be interpreted as a response of the developed fuzzy model rather than as a universal linear relationship.

The strong response to relative humidity also helps explain the differences between the surfaces shown in Figure 4. When RH was included as one of the varying variables, the model was able to produce substantially higher rainfall predictions under high-humidity conditions. In contrast, when RH was fixed in the temperature–evaporation surface, the variation in predicted rainfall associated with temperature and evaporation alone was comparatively limited. Therefore, the results suggest that relative humidity plays an important role in modifying the effects of temperature and evaporation within the FIS-RH-T-E. However, this interpretation represents the response of the developed model and should not be considered evidence that relative humidity independently controls rainfall, since rainfall formation results from the interaction of several atmospheric processes.

3.6 Performance of the FIS and the machine-learning models

Table 3 and Figure 8 compare the FIS-RH-T-E with the data-driven models. Climatology alone achieved NSE = 0.45 (R^2 = 0.45). The FIS reached R^2 = 0.63, NSE = 0.63 (95% CI 0.57–0.70) and KGE = 0.71, with RMSE = 40.4 mm and PBIAS = −3.2%. Under leave-one-year-out validation, MLR was the best data-driven model (R^2 = 0.67, NSE = 0.66, KGE = 0.72, RMSE = 38.7 mm), followed by SVR, GBR and RF (R^2 0.61–0.63, NSE 0.61–0.62, RMSE 41.4–41.8 mm), whereas the MLP was weakest (R^2 = 0.48, NSE = 0.44). The FIS therefore ranked second of the seven models.

Fitted to all 180 months, Eq. (5) became $\hat{P} = -231.6 + 4.70\ RH + 1.83\ T + 4.24\ E$, where $\hat{P}$ is the monthly rainfall total (mm month^{-1}), RH and T are monthly means (% and °C) and E is the monthly mean of daily pan evaporation (mm d^{-1}). Across the 15 leave-one-year-out fits, the RH coefficient varied only between 4.36 and 4.99, consequently each additional percentage point of RH adds about 4.7 mm of monthly rainfall. The positive coefficients of T and E, despite their negative simple correlations with rainfall, reflect multicollinearity: once RH is included, T and E contribute only small corrections, consistent with their weak partial correlations (Table, 2).

[Insert Table 3 here]

Table 3 Performance of the FIS-RH-T-E and the data-driven models for monthly rainfall estimation. LOYO: leave-one-year-out; MSESS: skill score against climatology; last two columns: 2020–2024 hold-out and in-sample (all 180 months). The FIS is identical in all schemes.

Model	R^2	NSE (95% CI)	KGE	RMSE (mm)	PBIAS (%)	MSESS (95% CI)	NSE 2020–2024	NSE in-sample
Climatology	0.45	0.45 (0.34–0.54)	0.57	49.6	0.0	0	0.43	0.52
FIS-RH-T-E	0.63	0.63 (0.57–0.70)	0.71	40.4	−3.2	0.34 (0.13–0.51)	0.57	0.63
MLR	0.67	0.66 (0.61–0.71)	0.72	38.7	3.2	0.39 (0.22–0.52)	0.60	0.68
RF	0.61	0.61 (0.53–0.68)	0.73	41.8	0.0	0.29 (0.13–0.41)	0.54	0.78
GBR	0.62	0.62 (0.55–0.68)	0.69	41.4	−1.7	0.30 (0.15–0.43)	0.55	0.76
SVR	0.63	0.62 (0.56–0.67)	0.64	41.4	−9.1	0.30 (0.14–0.43)	0.57	0.68
MLP	0.48	0.44 (0.29–0.56)	0.67	49.8	0.5	−0.01 (−0.32–0.21)	0.36	0.84

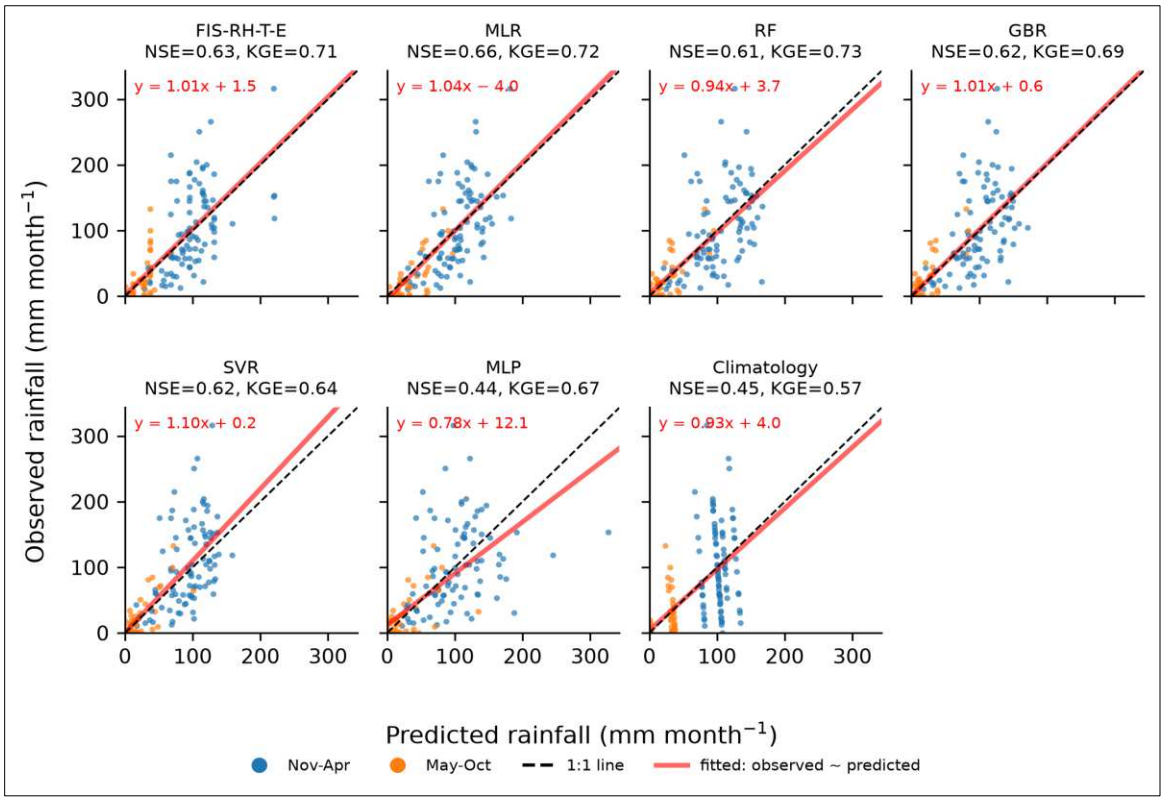


Fig. 8 Observed versus predicted monthly rainfall for the FIS-RH-T-E and the data-driven models. Data-driven models and climatology: leave-one-year-out predictions. Dashed line: 1:1; red line: regression of observed on predicted; blue: November–April; orange: May–October.

Fitted and scored on all 180 months, the data-driven models appeared considerably better NSE 0.68–0.84 (Table, 3), the MLP most of all (0.84). The difference between the in-sample and leave-one-year-out scores, up to 0.40 NSE units for the MLP, shows how strongly performance on the training data overstates accuracy for new years. Because the FIS has no fitted parameters but was designed with knowledge of the whole record, its scores should be read against both references: in-sample, the FIS (NSE = 0.63) was below RF, GBR and the MLP (0.76–0.84) and close to MLR and SVR (0.68); under leave-one-year-out validation, it was exceeded only by MLR. The paired bootstrap showed that the RMSE of the FIS did not differ significantly from that of MLR (+1.6 mm, 95% CI −0.4 to 3.5), RF (−1.4 mm, −6.6 to 3.1), GBR (−0.9 mm, −6.0 to 3.6) or SVR (−1.0 mm, −5.3 to 2.4), but was clearly lower than that of the MLP (−9.6 mm, −15.2 to −4.3) and of climatology (−9.0 mm, −16.0 to −2.9).

Relative to climatology (Figure 9a), MSESS was 0.39 for MLR, 0.34 for the FIS (95% CI 0.13–0.51) and 0.29–0.30 for RF, GBR and SVR; the MLP did not beat climatology (MSESS = −0.01). In the wet season (November–April),

anomaly correlations were 0.62 for MLR, 0.58 for the FIS and 0.52–0.56 for RF, GBR and SVR, i.e. the models captured part of the month-to-month variation beyond the seasonal mean. On the 2020–2024 hold-out, MLR (NSE = 0.60), SVR (0.57) and the FIS (0.57) led. The FIS had the smallest bias of these models in this period (PBIAS = +7.7%, compared with +9 to +24% for MLR, SVR, GBR and RF), which includes the dry years 2021 (292 mm) and 2022 (479 mm).

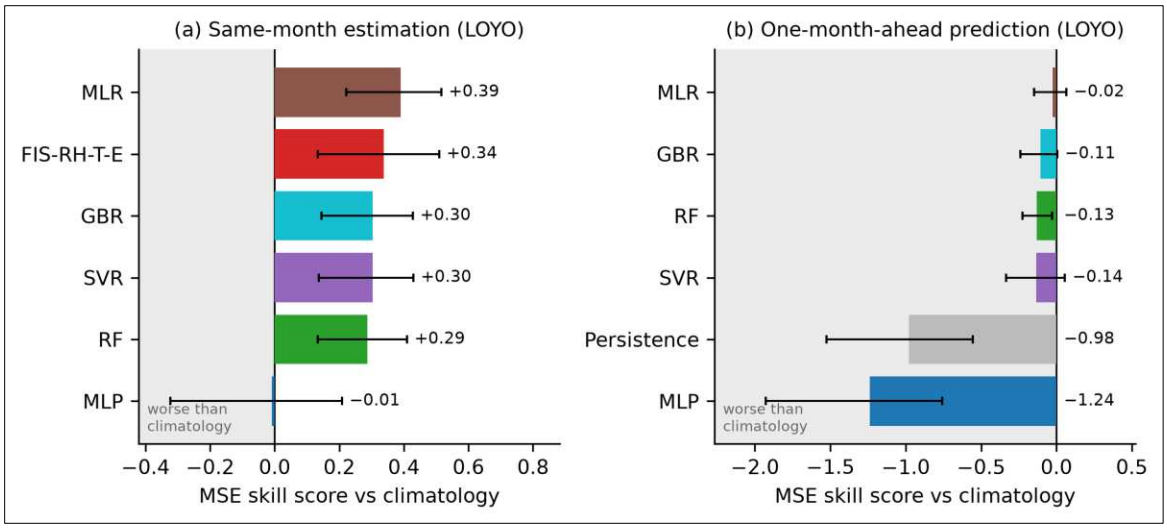


Fig. 9 Mean-squared-error skill score against climatology with 95% year-block bootstrap intervals for (a) concurrent estimation and (b) one-month-ahead prediction. Numbers give the skill score of each model; grey area: worse than climatology. Note the different scales of the x-axes.

3.7 Temporal pattern and high-rainfall months

Figure 10 compares the observed and predicted series. All models reproduced the seasonal cycle and most wet and dry months, but underestimated the largest events. The FIS predicted at most 221 mm, whereas observed rainfall reached 316.1 mm (December 2018). For the upper decile of months (≥ 157 mm; 18 months with a mean of 199 mm), the FIS predicted 55% of the observed mean and the data-driven models 51–58% (Figure 11). The FIS produced the widest range of predictions, but its highest values did not always coincide with the wettest months: its maximum of 221 mm was predicted for December 2014, when 118 mm were observed.

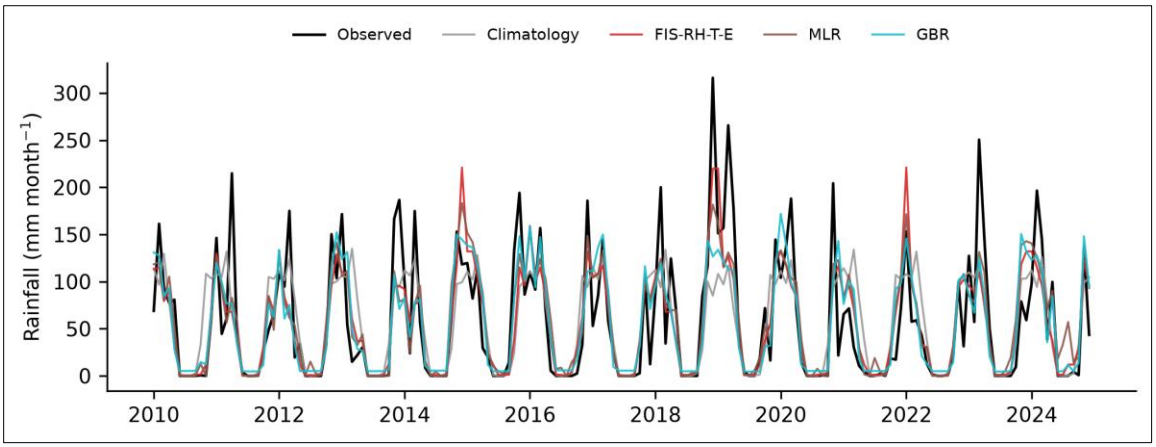


Fig. 10 Observed and predicted monthly rainfall, 2010–2024: climatology, FIS-RH-T-E, MLR and gradient boosting (GBR).

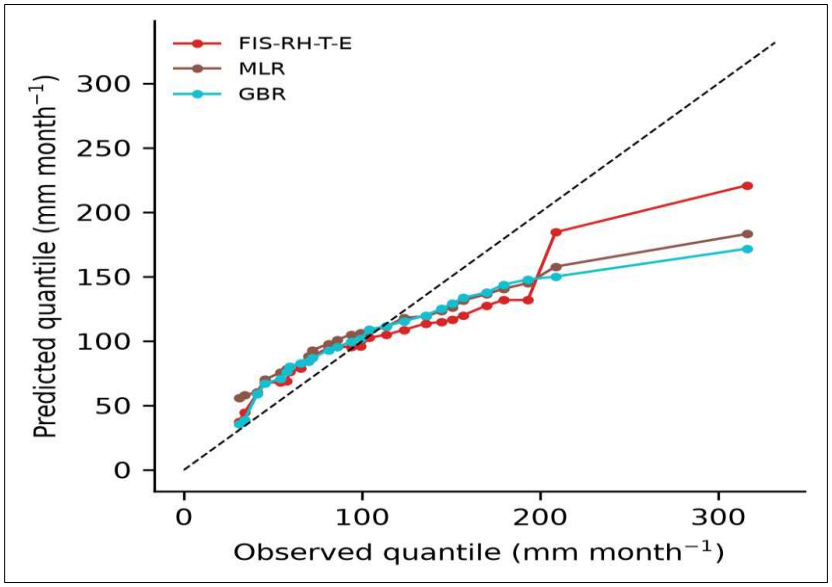


Fig. 11 Observed versus predicted upper quantiles of monthly rainfall for the FIS-RH-T-E, MLR and gradient boosting (GBR).

Two mechanisms are involved. First, in a Mamdani FIS with centroid defuzzification the crisp output always lies inside the output universe, below its upper limit of 250 mm, therefore a month such as December 2018 cannot be reproduced. Second, every model that is designed or fitted to minimise errors over many ordinary months is pulled towards typical values in the few extreme ones; this regression to the mean affected all data-driven models as well.

Widening the output universe of the FIS would remove the first limitation but not the second. Quantile post-processing or weighting of rare wet months during model development are possible remedies.

3.8 One-month-ahead prediction

Replacing concurrent inputs with those of the previous month removed all skill (Table, 4 and Figure 9b). No model beat climatology: the best, MLR, reached NSE = 0.44 and MSESS = −0.02 (95% CI −0.15 to 0.07), compared with NSE = 0.45 for climatology and −0.09 for persistence. RF, GBR, SVR and the MLP all had negative skill scores, and their anomaly correlations were close to zero (−0.11 to 0.14). The station variables therefore carry no useful information about next month's rainfall beyond the seasonal cycle, which suggests that part of the concurrent skill reflects the effect of rainy days on humidity and evaporation rather than a precursor of rain.

[Insert Table 4 here]

Table 4 One-month-ahead prediction with lagged inputs (leave-one-year-out; data-driven models only).

Model	NSE	KGE	RMSE (mm)	MSESS (95% CI)	p(MSESS ≤ 0)
Climatology	0.45	0.57	49.6	0	–
Persistence	−0.09	0.46	69.8	−0.98 (−1.52 to −0.55)	1.00
MLR	0.44	0.54	50.2	−0.02 (−0.15 to 0.07)	0.67
RF	0.38	0.49	52.7	−0.13 (−0.23 to −0.03)	0.99
GBR	0.39	0.51	52.2	−0.11 (−0.24 to 0.01)	0.97
SVR	0.38	0.38	52.9	−0.14 (−0.33 to 0.06)	0.91
MLP	−0.23	0.46	74.2	−1.24 (−1.93 to −0.76)	1.00

3.9 Implications and limitations

Because RH, T, E and rainfall share one dominant annual cycle, climatology alone explained 45% of the variance of monthly rainfall, hence high R^2 values of seasonal rainfall models mainly reflect the annual cycle. The FIS and best data-driven models reduced the remaining squared error by 29–39%, which is the skill attributable to the meteorological inputs; reporting skill against climatology should therefore become standard for fuzzy and machine-

learning rainfall models. That MLR outperformed RF, GBR, SVR and the MLP is consistent with the data: 180 monthly records, three strongly collinear inputs (VIF 9–12) and one dominant predictor whose relationship with rainfall anomalies is close to linear. Flexible models have little to gain and more variance to lose in this setting, and the MLP overfitted most clearly.

For practical use, FIS and MLR are suited to filling gaps and extending rainfall records at stations that measure T, RH and E; MLR estimated monthly rainfall with an RMSE of 38.7 mm (68% of the mean) and FIS with 40.4 mm (70%). The FIS has the additional advantage that its reasoning is expressed as readable rules that can be checked against physical understanding (Rudin 2019). Neither approach can forecast next month's rainfall from station data; operational forecasts would require predictors such as large-scale circulation indices or bias-corrected seasonal forecasts, which could be added to the FIS as further linguistic inputs.

The study has limitations. It uses one station, 15 years and monthly totals, which limits the number of wet extremes and prevents conclusions about spatial transferability. The membership functions and rules of FIS were defined with knowledge of the whole record, thus its scores are not fully independent of the evaluation data. All data-driven models overestimated rainfall in 2020–2024, suggesting that relationships learned in wetter years transfer imperfectly to drought conditions. Pan evaporation is sensitive to instrument exposure, and the unusually high temperatures recorded in summer 2012 should be confirmed. Optimising the membership functions and rules, for example with adaptive neuro-fuzzy inference (Sojitra et al. 2015), and testing the models at several stations are natural next steps.

Conclusion

A Mamdani fuzzy inference system was built in MATLAB with relative humidity, temperature, and evaporation as input variables, which then compared with five machine learning and statistical models for estimating monthly rainfall at Sulaymaniyah (2010–2024). Four conclusions follow: (1) Most of the apparent accuracy of monthly rainfall models comes from the seasonal cycle: climatology alone reached NSE = 0.45 and the meteorological inputs reduced the remaining squared error by up to 39%. Skill scores against climatology, not R^2 alone, should be reported. (2) Evaluated on the same months, the FIS reached R^2 = 0.63 and NSE = 0.63, second only to multiple linear regression (R^2 = 0.67, NSE = 0.66) and ahead of random forest, gradient boosting and support vector regression (NSE = 0.61–0.62); the differences among these five models were not statistically significant. A transparent, expert-

556 defined rule base can therefore match machine learning accuracy for this task, while expressing its reasoning in
557 readable rules.
558 (3) Relative humidity dominated both the FIS response surfaces and the data-driven analyses, in line with the
559 Mediterranean origin of the moisture that produces rainfall in the region.
560 (4) All models underestimated the wettest months, and none predicted rainfall one month ahead better than
561 climatology. The models are therefore suited to gap-filling and record extension rather than forecasting; future work
562 should optimize the membership functions and rules, add large-scale predictors and test the approach at several
563 stations.
564
565 **References**

566 Abdulrahman AM (2023) An analytical study of the climatic elements in Bazian region-Sulaymaniyah for the period
567 from 1985-2022 and its impact on the planting dates of some crops. Kirkuk University Journal for Agricultural
568 Sciences (KUJAS) 14:399-412. https://doi.org/10.58928/ku23.14339
569 Akgun A, Sezer EA, Nefeslioglu HA, Gokceoglu C, Pradhan B (2012) An easy-to-use MATLAB program
570 (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Computers &
571 Geosciences 38:23-34. https://doi.org/10.1016/j.cageo.2011.04.012
572 Al-Naseri SK, Falih AH, Kumar US (2022) Moisture sources and spatio-temporal variation of isotopic signatures in
573 Iraqi precipitation. Environmental Earth Sciences 81:435. https://link.springer.com/article/10.1007/s12665-022-
574 10559-7
575 Asklany SA, Elhelow K, Youssef IK, Abd El-wahab M (2011) Rainfall events prediction using rule-based fuzzy
576 inference system. Atmospheric Research 101:228-236.
577 https://www.sciencedirect.com/science/article/abs/pii/S0169809511000640
578 Breiman L (2001) Random forests. Machine Learnings 45:5–32.
579 https://link.springer.com/article/10.1023/A:1010933404324
580 Damayanti DR, Wicaksono S, Al Hakim MF, Jumanto J, Subhan S, Ifriza YN (2022) Rainfall prediction in Blora
581 regency using mamdani's fuzzy inference system. Journal of Soft Computing Exploration 3:62-69.
582 https://shmpublisher.com/index.php/joscex/sv/article/view/69

583 Elsner JB, Schmertmann CP (1994) Assessing forecast skill through cross validation. Weather and Forecasting
584 9:619-624. https://doi.org/10.1175/1520-0434(1994)009<0619:AFSTCV>2.0.CO;2
585 Fallah-Ghalhary GA, Mousavi-Baygi M, Habibi Nokhandan M (2009) Annual rainfall forecasting by using Mamdani
586 fuzzy inference system. Research Journal of Environmental Sciences 3:400-
587 413. https://scialert.net/abstract/?doi=rjes.2009.400.413
588 Fattah MA, Aziz SA, Kader MS (2020) Mitigating drought impact in corn cultivation using different drip irrigation
589 techniques. Sulaimani Journal for Pure and Applied Sciences 22:185-196.
590 https://sjpas.univsul.edu.iq/article?id=109
591 Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of Statistics 29:1189–
592 1232. https://www.jstor.org/stable/2699986
593 Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE
594 performance criteria: implications for improving hydrological modelling. Journal of Hydrology 377:80-91.
595 https://doi.org/10.1016/j.jhydrol.2009.08.003
596 Hamaamin RA, Ali OMA, Kareem SW (2025) Climate Forecasting in Sulaymaniyah City Using Deep Learning
597 Techniques. Sulaimani Journal for Engineering Sciences 11:54-74. https://doi.org/10.17656/sjes.10196
598 Hamaamin YA (2025) Evaluation of data mining and artificial intelligence methods to predict daily
599 precipitation. Journal of Engineering 31:99-112. https://doi.org/10.31026/j.eng.2025.05.06
600 Hamabor NO, Fattah MA, Khwakaram AI (2023). Fuzzy inference system in assessing the improved wastewater
601 quality for irrigation through sedimentation, dilution, and biochar filtration in Sulaymaniyah city, Iraq. IOP
602 Conference Series: Earth and Environmental Science 1214:012007. https://doi.org/10.1088/1755-
603 1315/1214/1/012007
604 Harris CR, Millman KJ et al (2020) Array programming with NumPy. Nature 585:357-362.
605 https://doi.org/10.1038/s41586-020-2649-2
606 Hasan M, Tsegaye T, Shi X, Schaefer G, Taylor G (2008) Model for predicting rainfall by fuzzy set theory using
607 USDA scan data. Agricultural Water Management 95:1350-1360.
608 https://doi.org/10.1016/J.AGWAT.2008.07.015
609 Hunter JD (2007) Matplotlib: A 2D graphics environment. Computing in Science & Engineering 9:90-95.
610 https://doi.org/10.1109/MCSE.2007.55

611 Janarthanan R, Balamurali R, Annapoorani A, Vimala V (2021) Prediction of rainfall using fuzzy logic. Materials
612 Today: Proceedings 37:959-963. https://doi.org/10.1016/j.matpr.2020.06.179
613 Jantakoon N (2016). Statistics Model for Meteorological Forecasting Using Fuzzy Logic Model. Mathematics and
614 Statistics 4:95-100. http://dx.doi.org/10.13189/ms.2016.040401
615 Katambara Z, Ndiritu J (2009) A fuzzy inference system for modelling streamflow: case of Letaba River, South
616 Africa. Physics and Chemistry of the Earth, Parts A/B/C 34:688-700. https://doi.org/10.1016/j.pce.2009.06.001
617 Małolepsza O, Mikołajewski D, Prokopowicz P (2025) Using fuzzy logic to analyse weather conditions. Electronics
618 14:85.
619 https://doi.org/10.3390/electronics14010085?urlappend=%3Futm_source%3Dresearchgate.net%26utm_mediu
620 m%3Darticle
621 McKinney W (2010) Data structures for statistical computing in Python. In: van der Walt, S. and Millman, J. (Eds.),
622 Proceedings of the 9th Python in Science Conference 56-61. https://doi.org/10.25080/Majora-92bf1922-00a
623 Michaelsen J (1987) Cross-validation in statistical climate forecast models. Journal of Climate and Applied
624 Meteorology 26:1589–1600. https://doi.org/10.1175/1520-0450(1987)026%3C1589:CVISCF%3E2.0.CO;2
625 Mohammadzadeh H, Eskandari Mayvan J, Heydarizad M (2020) The effects of moisture sources and local
626 parameters on the ^{18}O and ^{2}H contents of precipitation in the west of Iran and the east of Iraq. Tellus B:
627 Chemical and Physical Meteorology 72:1721224. https://doi.org/10.1080/16000889.2020.1721224
628 Mohd Safar NZ, Ramli AA, Mahdin H, Ndzi D, Khalif KMNK (2019). Rain prediction using fuzzy rule based
629 system in North-West Malaysia. Indonesian Journal of Electrical Engineering and Computer Science 14:1572-
630 1581. https://doi.org/10.11591/ijeecs.v14.i3.pp1572-1581
631 Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for
632 systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50:885-900.
633 https://doi.org/10.13031/2013.23153
634 Murphy AH (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient.
635 Monthly Weather Review 116:2417-2424. https://doi.org/10.1175/1520-
636 0493(1988)116%3C2417:SSBOTM%3E2.0.CO;2
637 Najmaddin PM, Whelan MJ, Balzter H (2017) Estimating daily reference evapotranspiration in a semi-arid region
638 using remote sensing data. Remote Sensing 9:779. https://doi.org/10.3390/rs9080779

Pedregosa F, Varoquaux G et al (2011) Scikit-learn: machine learning in Python. Journal of Machine Learning Research 12:2825-2830. https://jmlr.org/papers/v12/pedregosa11a.html

Piñeiro G, Perelman S, Guerschman JP, Paruelo JM (2008) How to evaluate models: observed vs. predicted or predicted vs. observed?. Ecological Modelling 216:316-322. https://doi.org/10.1016/j.ecolmodel.2008.05.006

Rahman MA (2020) Improvement of rainfall prediction model by using fuzzy logic. American Journal of Climate Change 9:391. https://doi.org/10.4236/ajcc.2020.94024

Roberts DR, Bahn V et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40:913-929. https://doi.org/10.1111/ecog.02881

Ross TJ (2009) Fuzzy logic with engineering applications. John Wiley & Sons.

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1:206-215. https://doi.org/10.1038/s42256-019-0048-x

Selvaraj A, Arasan A et al (2023) Mamdani Fuzzy Based Prediction of Rainfall Fuzzy Rainfall Index. In International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022):838-850. https://doi.org/10.2991/978-94-6463-136-4_74

Sham FAF, El-Shafie A, Jaafar WZW, Sherif M, Ahmed AN (2025) Advances in AI-based rainfall forecasting: A comprehensive review of past, present, and future directions with intelligent data fusion and climate change models. Results in Engineering 27:105774. https://doi.org/10.1016/j.rineng.2025.105774

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Statistics and Computing 14:199-222. https://doi.org/10.1023/B%3ASTCO.0000035301.49549.88

Sojitra MA, Purohit RC, Pandya PA (2015) Comparative study of daily rainfall forecasting models using adaptive-neuro fuzzy inference system (ANFIS). Current World Environment 10:529. http://dx.doi.org/10.12944/CWE.10.2.19

The MathWorks Inc (2007) MATLAB R2007b and Fuzzy Logic Toolbox User's Guide *Version 2)*. The MathWorks Inc., Natick, Massachusetts, USA.

Tosun M, Dincer K, Baskaya S (2011) Rule-based Mamdani-type fuzzy modelling of thermal performance of multi-layer precast concrete panels used in residential buildings in Turkey. Expert Systems with Applications 38:5553-5560. https://doi.org/10.1016/j.eswa.2010.10.081

Van Rossum G, Drake FL (2009) *Python 3 Reference Manual*. CreateSpace, Scotts Valley, California.

Virtanen P, Gommers R et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17:261-272. https://doi.org/10.1038/s41592-019-0686-2

Zadeh LA (1965) Fuzzy sets. Information and control 8:338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

Competing Interests: The authors of this document collectively declare that there are no conflicts of interest regarding the content presented herein. This encompasses any financial or personal relationships with individuals or organizations that could potentially influence the objectivity or integrity of the information provided.

Funding: This research received no external funding

Author Contributions:

Shagul F. Karim: Writing-original draft; Mohammed A. Fattah: Conceptualization, Supervision, Writing-review and Editing. Hadi H. Muhammed: Methodology, Software, Formal analysis, Visualization, Writing-review and Editing. The remaining authors: Investigation, Data collection, Data cleaning, and Data organization.

Data availability: The data used in this study are available upon request from the corresponding author.