



Forecasting Monthly Pan Evaporation in Al-Najaf, Iraq Using SARIMA and Month-Wise ARMA Models for Water-Resource Risk Assessment

Alaa B. Hameed^{1*}, Mohammed S. Shamkhi^{1,2}

¹ Department of Structures and Water Resources, Faculty of Engineering, University of Kufa, Najaf 54001, Iraq

² Civil Engineering Department, College of Engineering, University of Warith Al-Anbiyaa, Kerbala 56001, Iraq

Corresponding Author Email: alaab.alibadi@student.uokufa.edu.iq

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijdne.210920>

ABSTRACT

Received: 2 July 2026

Revised: 24 August 2026

Accepted: 11 September 2026

Available online: 30 September 2026

Keywords:

pan evaporation, stochastic forecasting, water-resource risk assessment, SARIMA, month-wise Autoregressive Moving Average, arid regions

Predicting monthly pan evaporation is crucial for water-resource risk assessment in arid regions, particularly in Al-Najaf, Iraq, where hyper-arid conditions stress limited water reserves. The main scientific contribution of this study lies in establishing a rigorous, long-term comparative evaluation between constant-parameter seasonal stochastic structures and seasonally varying month-specific models, providing a decadal probabilistic baseline for regional evaporation variability. Monthly pan evaporation observations from Al-Najaf meteorological station for 1968–2024 (684 months) were chronologically split into 600 training months (1968–2017) and 84 out-of-sample validation months (2018–2024). Utilizing the Box-Jenkins methodology, a constrained Seasonal Autoregressive Integrated Moving Average (SARIMA) $(0,1,[1,2,4])(0,1,1)_{12}$ model and twelve month-specific Autoregressive Moving Average (ARMA) models were developed following Box-Cox variance stabilization, stationarity verification, and residual diagnostics. During the 2018–2024 validation period, both approaches demonstrated statistical convergence with an identical Nash-Sutcliffe efficiency (NSE = 0.671) and comparable root mean square errors (RMSE = 62.40 mm for SARIMA vs. 62.21 mm for month-wise ARMA; Diebold-Mariano (DM) $p > 0.05$). While month-wise ARMA captures individual seasonal persistence, SARIMA achieved lower mean absolute errors (MAE = 47.80 mm vs. 48.46 mm), lower mean absolute percentage errors (MAPE = 27.4% vs. 30.5%), and greater parsimony, motivating its selection as the reference forecasting baseline. The resulting 2025–2034 projection preserves the historical seasonal cycle and establishes a decadal statistical reference for atmospheric evaporative variability, supporting regional risk assessment rather than serving as a direct deterministic measure of future water losses.

1. INTRODUCTION

Evaporation is a major component of the hydrological cycle and a substantial pathway of water loss in arid and semi-arid regions. In Al-Najaf city, Iraq, high evaporative demand occurs in a water-scarce setting where population growth and climatic variability increase pressure on limited water resources [1, 2].

From an ecodynamics perspective, atmospheric evaporative demand in hyper-arid environments acts as a primary regulatory driver of ecosystem water balance, soil moisture depletion and land degradation across fragile arid soils governed by textural retention properties [3, 4]. In vulnerable natural systems such as the Bahr Al-Najaf wetland depression, extreme summer evaporative rates drive severe seasonal lake contraction and surface water shrinkage [5], while extended regional drought intensifies overall eco-hydrological vulnerability [6]. With accelerating climate variability across the Euphrates basin [7] and projected escalations in evaporative demand [2], establishing statistical baselines of monthly evaporation is therefore critical for both water-

resource risk evaluation and sustainable environmental assessment in such arid settings.

Accurate evaporation forecasting is essential for strategic planning in dam management and irrigation projects [8]. Physical and empirical approaches for estimating evaporative losses commonly require several meteorological inputs. Their performance can deteriorate when input observations are sparse, inconsistent, or locally unrepresentative, which motivates complementary data-driven forecasting of directly observed pan evaporation [9, 10]. Recent research suggests that these conventional models may struggle to represent the complex and local dynamics of evaporation [8]. This has led researchers to adopt advanced statistical tools for hydrological forecasting, as Machiwal and Jha [11] illustrated in their book on theoretical and applied hydrological modeling.

The methodologies utilized in modeling climate variables vary, but among them, the Autoregressive Integrated Moving Average (ARIMA) model and Seasonal Autoregressive Integrated Moving Average (SARIMA) have emerged as important tools in hydrology for modeling and predicting climate parameters [12]. These models are distinguished by

their ability to observe linear and seasonal relationships in climate data, leading to their widespread use in recent studies in various geographical regions, including Iraq [13]. A study [14] laid the foundation for this methodology, which relies on transforming non-stationary data into stationary series to ensure forecast accuracy.

While machine-learning and artificial intelligence models have increasingly been adopted for hydro-climatic forecasting in semi-arid regions [9, 15, 16], classical Box-Jenkins stochastic models remain fundamental benchmarks for long-term univariate evaporation series. Specifically, stochastic models offer structural parsimony, explicit mathematical transparency, and formally derived probabilistic prediction intervals, providing an interpretable statistical foundation when multivariable meteorological drivers are unavailable or unobserved [17, 18].

Regarding evaporation modeling, Sule and Ajala [19] demonstrated that the ARIMA methodology is superior in long-term forecasting and overcomes the challenges of missing data. This was also demonstrated in a study by Benchaiba et al. [20]. In Algeria, SARIMA models proved to be a more accurate alternative to physical methods based on the Dalton formula, especially in semi-arid regions where direct measurements lack precision.

Beyond evaporation, SARIMA has been applied to several hydro-climatic variables, including temperature, rainfall, river discharge, and drought indices, confirming the general suitability of seasonal Box-Jenkins models for strongly periodic environmental series. However, these applications do not resolve whether a constant-parameter seasonal model or a periodically varying parameter structure is more appropriate for long monthly pan evaporation records in Iraq [21–23].

However, given the climatic variations occurring in semi-arid regions, research has demonstrated that SARIMA models may not always be sufficient. Because it assumes constant parameters over time, researchers have turned to month-wise Autoregressive Moving Average (ARMA) models. These models allow statistical parameters to change periodically according to the season, thus simulating the behavior of each season separately [18]. A study by Shimi et al. [10] in Iran demonstrated that ARMA models offer superior predictive power compared to traditional linear models, due to their flexibility in representing periodic changes in statistical parameters. Similarly, Wang et al. [24] showed that constructing discrete ARMA models for individual monthly subsequences effectively captures interannual continuity, reducing the loss of month-specific variance inherent in conventional seasonal models.

Despite the growing use of stochastic models in hydro-climatology, evidence remains limited on how constant-parameter seasonal models compare with seasonally varying stochastic structures for long monthly pan evaporation records in Iraq. In particular, a transparent out-of-sample comparison using a long record from Al-Najaf, together with long-horizon baseline projections and explicit uncertainty bounds, is still needed [15, 25]. Accordingly, the objectives were to: (1) calibrate competing SARIMA and month-wise ARMA models using maximum likelihood estimation (MLE) on the 1968–2017 training record; (2) compare their out-of-sample forecast performance during 2018–2024 using root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and additional skill diagnostics; and (3) generate a 10-year baseline forecast for 2025–2034 with prediction intervals to support water resource planning. By

quantifying both seasonal dynamics and forecast uncertainty, the study is intended to provide a reproducible statistical baseline for evaluating evaporative water-loss risk in an arid urban region, thereby connecting dynamic environmental modeling with sustainable water-resource management.

2. METHODOLOGY

2.1 Study area and data collection

This study is conducted in Al-Najaf, located in the central-southern region of Iraq (Figure 1), which is characterized by a semi-arid climate. This region experiences extreme temperature variations and high evaporation rates, particularly during the summer months, which directly impacts the local hydrological balance [6]. This study is based on 57 years of historical monthly pan evaporation records, from 1968 to 2024, with a total of 684 observations. These long-term data are essential for understanding climatic trends and cyclical fluctuations, which are necessary for constructing accurate stochastic models [18]. The data were obtained from the Iraqi Meteorological Organization and Seismology (IMOS), Ministry of Transport, measured using a standard Class A evaporation pan located at Al-Najaf meteorological station (station ID: 40670, latitude: 31.9500°N, longitude: 44.3166°E, elevation: 50 m a.s.l.) and expressed in mm month⁻¹.

The total monthly record of 684 observations was divided into two sets: a training set of 600 months (January 1968 to December 2017) for model estimation, and an independent validation set of 84 months (January 2018 to December 2024) to evaluate predictive performance. A total of thirty-four monthly observations (5.6% of the 600-month training record) were missing, all occurring strictly within the training period as short, isolated gaps not exceeding two consecutive months due to routine gauge maintenance and record archiving transitions. An inter-annual seasonal interpolation scheme—interpolating missing values across corresponding months of adjacent years—was applied for data imputation [26, 27]. All imputation statistics were derived exclusively from the training period. The 2018–2024 validation period contained complete observations with zero missing values.

To ensure no systematic bias was introduced into the imputation method, statistical properties before and after imputation were verified; the series mean shifted negligibly from 284.65 mm to 286.02 mm (<0.5%), and the standard deviation shifted from 177.71 mm to 178.49 mm (<0.45%), with formal two-sample difference and variance testing confirming distributional preservation without statistical bias ($t = 0.131$, $p = 0.895$, variance ratio $F = 1.009$, $p = 0.913$). Furthermore, a formal sensitivity analysis was performed comparing the inter-annual scheme with the Kalman seasonal state-space smoothing method. The inter-annual scheme demonstrated superior accuracy, achieving lower error metrics (RMSE, MAE, symmetric mean absolute percentage error (sMAPE) and MAPE), while overall model rankings, dynamical structures, and validation inferences (SARIMA vs. month-wise ARMA) remained identical in both approaches. Prior to modeling, the record was screened for abrupt structural breaks and temporal heterogeneity. A homogeneity test across the 1968–2024 record (using Pettitt's test) showed a statistically significant shift ($p < 0.05$) centered around 2008. Official station data confirmed that the Najaf station maintained continuous operation at fixed geographic

coordinates and elevation using a standard Class A pan (120.7 cm diameter, 25.4 cm depth) under the uniform observational guidelines of the World Meteorological Organization (WMO) [28], confirming that this shift represents a genuine climatic signal resulting from regional drought trends, and not an observational anomaly. This statistical shift around 2008 is corroborated by documented hydro-climatic records in central Iraq and Al-Najaf station, where the severe regional drought of 2007–2009 triggered sharp precipitation deficits, marked positive temperature anomalies, and extreme negative drought indices (RDI and SPI) across the Middle Euphrates [29–31].

To account for this non-stationarity, first-order differencing and seasonal differencing ($d = 1, D = 1$) were applied within the SARIMA modeling framework, effectively turning the series into a stationary process before parameter estimation.

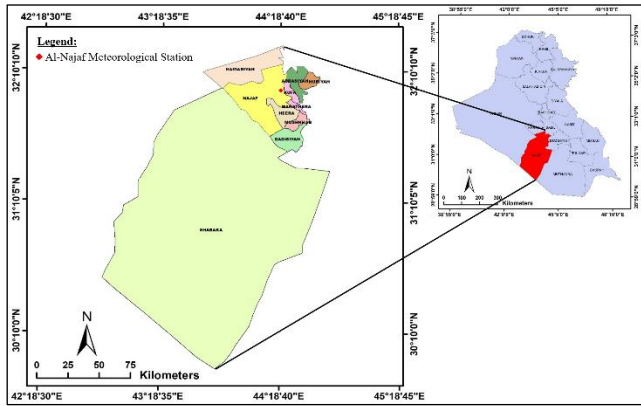


Figure 1. Geographic location of the study area (Al-Najaf, Iraq)

2.2 Data transformation and stationarity

Time series stationarity is a fundamental requirement for statistical models [32]. To obtain this, the original series of monthly pan evaporation depth was examined utilizing the Augmented Dickey-Fuller (ADF) test [33]. The ADF test was used to assess the presence of a unit root in the conditional mean process. Variance heterogeneity and distributional skewness were assessed separately using descriptive diagnostics and motivated the Box-Cox transformation.

The optimal Box-Cox transformation parameter was estimated by maximizing the log-likelihood function over a range of candidate (λ) values. To stabilize the variance, the standard piecewise Box-Cox transformation was applied (Eq. (1)) [34].

$$Y_t = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(X_t) & \text{if } \lambda = 0 \end{cases} \quad (1)$$

where, Y_t is the transformed value at time t , X_t is the observed monthly pan evaporation depth at time t and λ is the Box-Cox transformation parameter.

For the SARIMA (600 months) training series, likelihood maximization yielded $\lambda = 0.260$. The transformed series was then assessed for stationarity [35], followed by Seasonal and, if justified, regular differencing.

For the month-wise ARMA, transformation requirements were assessed separately using a likelihood-based Box-Cox estimation and distributional diagnostics. Only December required a material transformation, with $\lambda \approx 0.01$, close to the

logarithmic transformation; the other monthly subseries were retained on their original scale because the transformation did not materially improve variance stability or the residual diagnostics. Then the twelve series were standardized independently by Eq. (2) [10].

$$Z_{y,m} = \frac{X_{y,m}^* - \mu_m}{\sigma_m} \quad (2)$$

where, $Z_{y,m}$: monthly standardized series, $X_{y,m}^*$: the transformed (or untransformed, if no transformation was applied) monthly value, μ_m and σ_m : monthly mean and standard deviation estimated from the training period only.

To strictly prevent data leakage and guarantee genuine out-of-sample evaluation, all preprocessing parameters were derived exclusively from the 600-month training record (1968–2017). Specifically, the Box-Cox parameter ($\lambda = 0.260$ for SARIMA; $\lambda = 0.01$ for December ARMA) and the month-specific standardization moments (μ_m and σ_m) were estimated solely on training data. During the 2018–2024 validation period, out-of-sample observations and their generated forecasts were back-transformed and unstandardized using these fixed training parameters without any re-estimation.

2.3 Mathematical formulation

SARIMA model is known as SARIMA (p,d,q) (P,D,Q)_s, whereas (p,d,q) refers to the non-seasonal part of the original series, (P,D,Q) refers to the seasonal part of the series and s is the season; this can be written as in (Eq. (3)).

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (3)$$

where, p : the order of nonseasonal autoregressive (AR) model, q : the order of nonseasonal Moving average (MA) model, d : the number of regular differencing, P : the order of seasonal AR, D : number of seasonal differencing, Q : the order of seasonal MA, S : the length of season cycle, B : demonstrates the backward operators, Φ : seasonal AR coefficient, ϕ : non-seasonal AR coefficient, $(1-B^S)$: seasonal differencing, $(1-B)^d$: non-seasonal differencing, θ : non-seasonal MA coefficient, and Θ : seasonal MA coefficient [36].

For comparison with SARIMA, the monthly record was decomposed into 12 calendar-month subseries, and an independent ARMA (p,q) model was fitted to each subseries across years. This approach captures month-specific interannual persistence but does not model dependence between adjacent calendar months; the mathematical model for each calendar month m is expressed in (Eq. (4)) [18]:

$$Z_{m,y} = \sum_{j=1}^{pm} \phi_{m,j} * Z_{m,y-j} + \varepsilon_{m,y} - \sum_{j=1}^{qm} \theta_{m,j} \varepsilon_{m,y-j} \quad (4)$$

where, $Z_{m,y}$ is the standardized evaporation value for calendar month (m) in year (y); p_m and q_m denote the month-specific non-seasonal AR and MA orders, respectively; $\phi_{m,j}$ and $\theta_{m,j}$ represent the month-specific AR and MA parameters; and $\varepsilon_{m,y}$ is the independent white noise error term for month (m) in year (y).

The month-wise ARMA formulation is used here as a pragmatic seasonal-subseries benchmark rather than as a full periodic ARMA process. It captures interannual persistence within each calendar month but omits cross-month dependence; this limitation is explicitly considered when

interpreting the comparison with SARIMA.

2.4 Model selection criteria

Information criteria were utilized to ensure parsimony and avoid overfitting. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were utilized as shown in Eqs. (5) and (6):

$$AIC = -2\ln(L) + 2k \quad (5)$$

$$BIC = -2\ln(L) + k \ln(n) \quad (6)$$

where, L : the max likelihood function of the model, k : the number of estimated parameters in the model, and n : the number of observations in the model. The model that obtains the min value for both the AIC and BIC is considered the most suitable for representing the data [37].

2.5 Model identification

This is the first stage in the Box-Jenkins methodology, an overview of which is provided in Figure 2. Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) are used to identify candidate time series models and their parameter orders (p, d, q) [17]. ACF and PACF plots were inspected using approximate 95% reference bounds of $\pm 1.96/\sqrt{N}$, where N represents the number of observations [38].

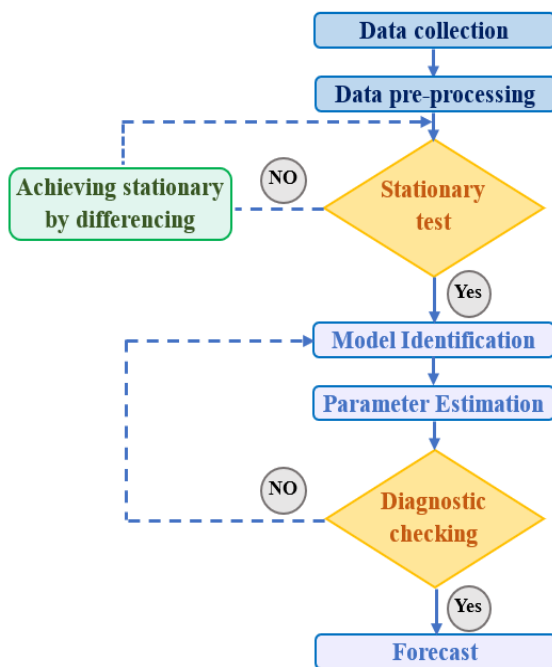


Figure 2. Box-Jenkins methodology plot

Table 1. Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) criteria for identifying the order of Autoregressive Moving Average (ARMA) model

Model	ACF	PACF
AR (p)	Decay towards zero	Spikes cutoff after lag p
MA (q)	Spikes cutoff after lag q	Decay towards zero
ARMA (p, q)	Decay towards zero	Decay towards zero

Candidate orders were selected from the observed cutoff/decay patterns and then compared by Information Criteria (AIC/BIC) to establish model optimum. Based on the statistical behavior of these functions, the methodological rules summarized in Table 1 were applied [19].

For the SARIMA framework, non-seasonal orders were systematically identified from correlation cutoff and decay patterns at early lags, while seasonal orders were inspected at seasonal cycle multiples ($S = 12$). To ensure a transparent and comprehensive model space, the search protocol evaluated candidate non-seasonal orders up to $p, q \leq 4$ and seasonal orders $P, Q \leq 1$ (with differencing fixed at $d = 1, D = 1$). For the month-wise ARMA framework, diagnostic identification was conducted independently on each calendar-month subseries across a parameter space bounded by $p, q \leq 2$, establishing candidate AR and MA configurations based on the theoretical rules in Table 1 [18, 24].

2.6 Parameter estimation

This is the second stage in the Box-Jenkins methodology and aims to find the parameter values for the models identified in the first step. In this study, the model parameters were estimated utilizing MLE available in the SPSS software.

For the SARIMA model, the estimation algorithms were applied to the entire series. Stationarity and invertibility were verified by confirming that the roots of the AR and MA characteristic polynomials satisfied the required conditions (all inverse characteristic roots lie strictly within the unit circle) [17, 23]. In the month-wise ARMA model, however, the estimation process was performed for each month separately, with the parameters for each monthly series being extracted independently. This allows each month to express its own statistical characteristics without being bound by the assumption of parameter stability across the year [18, 24].

To confirm the significance of these parameters, t -tests were conducted for each parameter; parameters with a p -value < 0.05 were considered statistically significant. To formally verify the parsimony principle when eliminating statistically insignificant parameters, the Likelihood Ratio (LR) test was applied to evaluate constrained subset models against their unrestricted counterparts [39] (Eq. (7)):

$$LR = -2(\ln L_{\text{restricted}} - \ln L_{\text{full}}) \sim \chi^2(df) \quad (7)$$

This parsimonious simplification ensured that there were no unnecessary coefficients that did not add explanatory value [11].

2.7 Diagnostic checking

After estimating the model parameters, the next stage is model diagnosis to assess the efficiency of data representation. This involves examining residuals, which are the differences between the actual values and the values predicted by the model. A model's success depends on making the residuals behave like a "white noise" process. This means they must be independent, have a zero mean, and a constant variance [16, 40].

Residual independence was evaluated with the Ljung-Box Q test at lag $h = 18$ using degrees of freedom (df) adjusted for the number of fitted AR/MA parameters [41]. Residual normality was examined to assess the adequacy of Gaussian prediction-interval assumptions, using Q-Q plots and formal

normality tests [42].

ACF and PACF were also plotted for the residuals, where the occurrence of all values within the confidence interval ($\pm 1.96/\sqrt{N}$) indicates the absence of statistically significant correlations between them [19].

2.8 Performance evaluation metrics

To assess the efficiency of the nominated models in predicting monthly pan evaporation depth and to identify the model best able to represent the data, the statistical metrics demonstrated below were utilized to assess the performance of the SARIMA and month-wise ARMA models. These metrics are essential evaluation tools in hydrological research, as they provide indicators of the difference between the actual values (Y_i) and the values predicted by the models ($\hat{Y}_i$) [23].

2.8.1 Root mean square error

Used to measure the overall error value, giving greater weight to large errors and penalizing extreme peak deviations, making it sensitive to outliers [43]. It is calculated utilizing Eq. (8):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (8)$$

2.8.2 Mean absolute error

Measures the average value of the actual errors on a continuous linear scale regardless of their direction, providing an unweighted baseline that avoids the quadratic inflation of RMSE [44]. It is calculated utilizing Eq. (9):

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (9)$$

2.8.3 Mean absolute percentage error

Provides a relative standard that facilitates comparison between models [9]. It is calculated utilizing Eq. (10):

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \right) \times 100 \quad (10)$$

2.8.4 Symmetric mean absolute percentage error

This measure assesses the percentage of relative error between observed and predicted values, giving equal weight to overestimates and underestimates [45]. It is calculated using Eq. (11):

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{\left(\frac{|Y_i| + |\hat{Y}_i|}{2} \right)} \times 100 \quad (11)$$

2.8.5 Nash-Sutcliffe efficiency

This measure assesses the predictive power and ability of hydrological models to accommodate variance compared to the observed mean variability [46]. It is calculated utilizing Eq. (12):

$$NSE = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (12)$$

2.8.6 Pearson's coefficient of determination

It measures the strength of linear correlation and the consistency of structural patterns between observed and predicted values [47]. It is calculated using Eq. (13):

$$Pearson R^2 = \frac{\sum (Y_i - \bar{Y})(\hat{Y}_i - \mu_{\hat{Y}})}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (\hat{Y}_i - \mu_{\hat{Y}})^2}} \quad (13)$$

where, $\bar{Y}$ is the mean of observed values, $\hat{Y}_i$ is the predicted value and $\mu_{\hat{Y}}$ is the mean of the predicted values. While Pearson R^2 quantifies collinearity and structural alignment [47], it is evaluated alongside NSE to provide a strict hydrological benchmark that penalizes variance differences and systematic bias [46]. These metrics are applied to ensure comprehensive model accuracy. A model with the lower error values (RMSE, MAE, MAPE, sMAPE), higher Pearson R^2 , and NSE is considered optimal [48].

To evaluate whether the differences in forecast accuracy between the competing models are statistically significant, the Diebold-Mariano (DM) test was applied to the validation forecast error series under both squared-error and absolute-error loss functions [49], as shown in Eq. (14):

$$DM = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{N}}} \quad (14)$$

where, $\bar{d}$: the mean loss differential ($d_t = e_{1,t}^2 - e_{2,t}^2$ for squared-error loss and $d_t = |e_{1,t}| - |e_{2,t}|$ for absolute error loss), s_d^2 : the sample variance of the loss differential series, $e_{1,t}$ and $e_{2,t}$ are the forecast errors of the competing models at time t , and N : the validation sample size.

All transformations, model orders, and parameters were strictly determined using the 1968–2017 training set only. Out-of-sample forecasts for the validation period (2018–2024) were generated using a fixed-origin multi-step forecasting protocol originating from December 2017. No future validation information was utilized at any stage of model selection, parameter estimation, or data preprocessing. Furthermore, forecasts generated on the Box-Cox transformed scale ($\lambda = 0.260$) were returned to the physical evaporation scale using the naive inverse Box-Cox transformation. Prediction limits were back-transformed consistently by applying the inverse transformation directly to the upper and lower bounds to preserve the specified coverage probability on the original physical scale prior to calculating evaluation metrics.

2.8.7 Seasonal naive benchmark model

To evaluate the practical forecasting skill and added value of the calibrated stochastic models, a seasonal naive approach was incorporated as a zero-parameter benchmark reference [50]. Consistent with the multi-step fixed-origin validation process originating from December 2017, the seasonal naive model projects future monthly pan evaporation by repeating the last observed annual seasonal cycle of the training series across the validation horizon ($\hat{Y}_i = Y_{i-12}$).

3. RESULTS AND DISCUSSION

3.1 Exploratory data analysis

The evaporation time series for Najaf (1968–2024) exhibits a pronounced annual seasonal cycle, consistent with strong intra-annual climatic variability in the study area. Descriptive data analysis, as demonstrated in Table 2, reveals significant variability, with a monthly mean of 273.31 mm and a standard deviation of 174.76 mm, indicating wide inter-seasonal variations. This variability, along with the skewness and kurtosis values, reflects the heterogeneous nature of the data, necessitating statistical transformations to prepare the series

for modeling [17]. The monthly evaporation behavior demonstrates variations in the statistical characteristics of each month; monthly climatology confirms a minimum mean evaporation in January (75.24) and a maximum in July (538.39). The standard deviation was also observed to be higher in the summer months and lower in the winter months (Table 2), confirming that the evaporation time series in the study area exhibits heteroscedasticity related to seasonality. The series graph (Figure 3) demonstrated a slight downward trend in the peak of evaporation over the last two decades, a pattern that requires statistical transformations.

Table 2. Descriptive statistics of the monthly pan evaporation depth series (mm) (1968–2024)

Month	Min	Max	Mean	Std. Deviation	Skewness	Kurtosis
January	38.20	114.70	75.24	18.32	-0.25	-0.77
February	57.60	165.00	107.14	27.23	0.13	-0.93
March	107.80	266.70	184.39	43.08	-0.09	-0.75
April	145.80	438.10	256.09	61.62	0.31	-0.08
May	189.80	562.40	357.46	85.69	0.09	-0.53
June	235.20	683.50	474.45	109.17	-0.41	-0.51
July	292.60	729.90	538.39	120.47	-0.54	-0.86
August	266.20	667.20	483.50	104.26	-0.36	-0.61
September	197.30	508.80	356.94	74.23	-0.39	-0.72
October	121.80	377.50	243.08	56.66	-0.12	-0.43
November	63.45	209.50	125.24	36.10	0.22	-0.53
December	37.10	154.30	77.86	24.25	0.64	0.52
Overall	37.10	729.90	273.31	174.76	0.58	-0.72

Table 3. Augmented Dickey-Fuller (ADF) unit root test results for the monthly pan evaporation depth (mm) across different transformation stages for Seasonal Autoregressive Integrated Moving Average (SARIMA) model

Series Stage	Test Specification	Lags (AIC)	5% Critical Value	ADF <i>t</i> -Statistic	<i>p</i> -Value	Conclusion
Original Data	Intercept	12	-2.86	-1.914	0.4886	Non-stationary
Box-Cox transforms	Intercept	12	-2.86	-2.082	0.4623	Non-stationary
Seasonal Differenced (<i>D</i> = 1)	None	12	-1.95	-7.502	<0.01	Stationary
Seasonal and Normal differenced (<i>D</i> = 1, <i>d</i> = 1)	None	12	-1.95	-9.206	<0.01	Stationary

Note: Lag lengths were selected automatically by AIC (maximum lag = 12). Akaike Information Criterion (AIC).

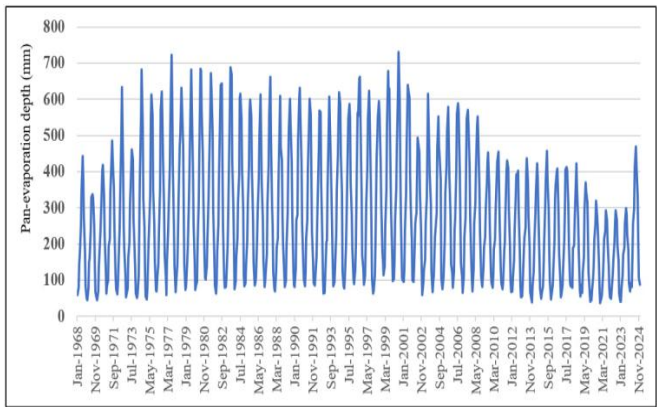


Figure 3. Time series plot of the monthly pan evaporation depth (mm) (1968–2024)

3.2 Model identification

Model identification is the first step in the Box-Jenkins methodology. It involves analyzing the statistical properties of

the series after processing. For the SARIMA model, the Box-Cox transformation ($\lambda = 0.260$) was applied to the entire time series to ensure variance stability. To evaluate stationarity, the ADF test was performed across all transformation stages, with lag lengths automatically selected using AIC (maximum lag = 12). An intercept specification was used for the original and Box-Cox transformed series, both of which failed to reject the unit-root null hypothesis ($t = -1.914, p = 0.4886$ and $t = -2.082, p = 0.4623$, respectively). Applying seasonal differencing at lag 12 ($D = 1$) successfully brought the series to stationarity ($t = -7.502, p < 0.01$, under a zero-mean specification). Regular differencing ($d = 1$) was additionally applied to remove residual short-term autocorrelation observed at early lags, further strengthening stationarity ($t = -9.206, p < 0.01$) without over-differencing the series. The complete unit-root test statistics across all stages are presented in Table 3. The ACF and PACF plots of the seasonal and normal differenced series ($d = 1, D = 1$) were examined up to 48 lags (Figures 4(a) and (b)). The plots demonstrated a cut-off pattern in the ACF plot and a gradual decay in the PACF plot,

suggesting non-seasonal MA components. The seasonal component demonstrated a distinct spike at lag 12 in the ACF plot, confirming the presence of a seasonal MA term. Based on the parsimony principle, four competitive candidate models were shortlisted: $(0,1,1)(0,1,1)_{12}$, $(0,1,2)(0,1,1)_{12}$, $(0,1,3)(0,1,1)_{12}$, and $(0,1,4)(0,1,1)_{12}$. These models will undergo parameter estimation and diagnostic testing to confirm their efficiency.

For the month-wise ARMA models, given the seasonal variation in monthly pan evaporation depth, each of the twelve months of the year was modeled as an independent time series.

This methodology requires precise identification of each monthly series. After standardizing all monthly series (by utilizing Eq. (2)) and applying the Box-Cox transformation to the December data ($\lambda = 0.01$), ACF and PACF plots were drawn for each month.

The correlation plots revealed that the cyclical pattern of evaporation is not constant across the months of the year, necessitating the selection of seven possible models: $(1,0)$, $(0,1)$, $(1,1)$, $(2,0)$, $(0,2)$, $(2,1)$, and $(1,2)$. These models were assessed based on the AIC and BIC criteria for each month.

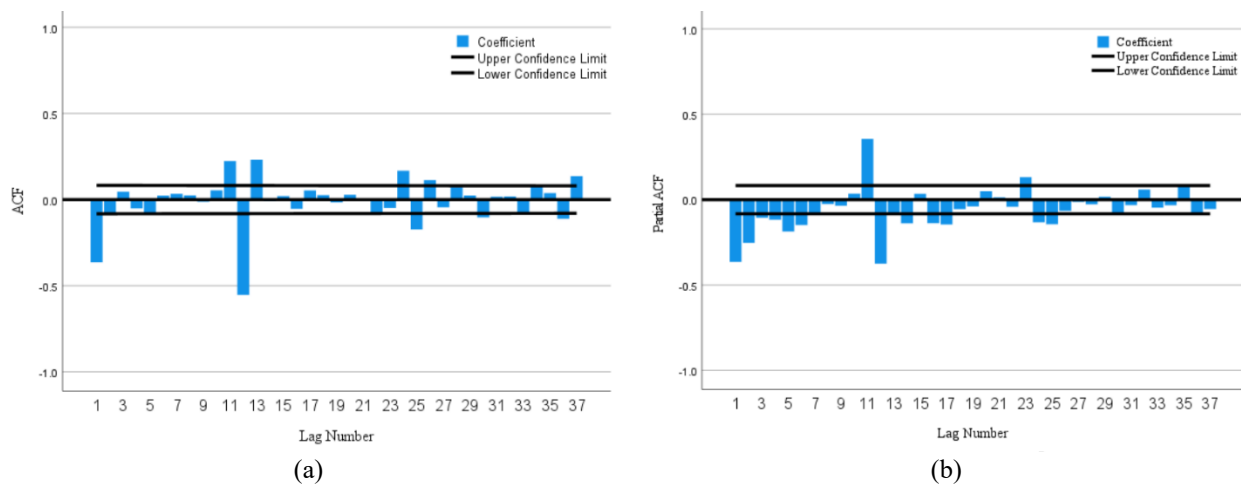


Figure 4. (a) Autocorrelation Functions (ACF) and (b) Partial Autocorrelation Functions (PACF) plots for Seasonal Autoregressive Integrated Moving Average (SARIMA) model after normal and seasonal differencing

Table 4. Parameter values of candidate Seasonal Autoregressive Integrated Moving Average (SARIMA) models for monthly pan evaporation depth (mm)

SARIMA Models	MA Parameters				Seasonal MA Parameters			
	θ_1		θ_2		θ_3		θ_4	
	Value	Sig.	Value	Sig.	Value	Sig.	Value	Sig.
$(0,1,1)(0,1,1)_{12}$	0.64	<0.001	-	-	-	-	-	-
$(0,1,2)(0,1,1)_{12}$	0.54	<0.001	0.24	<0.001	-	-	-	-
$(0,1,3)(0,1,1)_{12}$	0.54	<0.001	0.20	<0.001	0.057	0.17	-	-
$(0,1,4)(0,1,1)_{12}$	0.54	<0.001	0.15	0.002	-0.008	0.86	0.15	<0.001
$(0,1,[1,2,4])(0,1,1)_{12}$ constrained	0.54	<0.001	0.14	<0.001	-	-	0.14	<0.001

Table 5. Model selection criteria (Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)) for candidate Seasonal Autoregressive Integrated Moving Average (SARIMA) models for monthly pan evaporation depth (mm)

Models	AIC	BIC
$(0,1,1)(0,1,1)_{12}$	976.14	989.26
$(0,1,2)(0,1,1)_{12}$	947.34	964.84
$(0,1,3)(0,1,1)_{12}$	947.26	969.13
$(0,1,4)(0,1,1)_{12}$	937.32	963.57
$(0,1,[1,2,4])(0,1,1)_{12}$ constrained	935.36	957.23

Note: Bold text denotes the optimal selected model, which achieved the lowest AIC (935.36) and BIC (957.23) among all candidate models while retaining only statistically significant parameters ($p < 0.05$).

3.3 Parameter estimation

This is the second stage of the Box-Jenkins methodology, aiming to find the parameter values for each model. MLE algorithm in SPSS was utilized for this purpose.

For the SARIMA model, the parameters of the four candidate models were estimated (Table 4). When examining the statistical significance of the MA parameters (θ_1 , θ_2 , θ_3 , θ_4), it was found that the third parameter (θ_3) in the third and fourth candidate models was not statistically significant ($p > 0.05$).

Therefore, a new constrained model was constructed in which the insignificant parameter was excluded, while retaining the statistically significant fourth non-seasonal MA coefficient (θ_4), designated as the subset SARIMA $(0,1,[1,2,4])(0,1,1)_{12}$ model. Seasonal dependence was represented separately by the seasonal MA coefficient θ_1 at lag 12. To provide a statistical basis for removing the statistically insignificant θ_3 coefficient from the SARIMA $(0,1,4)(0,1,1)_{12}$ model, an LR test was conducted [39], which involved comparing the full (unrestricted) model with the constrained (restricted) model.

The log-likelihood for the full model ($\ln L = -462.66$) and for the constrained model ($\ln L = -462.68$) yielded an $LR = -2(\ln L_{\text{restricted}} - \ln L_{\text{full}}) = 0.04$ ($df = 1$, p -value = 0.850), compared to the chi-squared critical threshold at a significance level of 5% (i.e., 3.84), the test statistic fails to reject the null hypothesis confirming that excluding θ_3 does not degrade model performance.

Furthermore, the constrained model obtained the best results (lowest values) on both the AIC and the BIC, 935.36 and 957.23, respectively (Table 5). Therefore, this model was nominated as the best model for representing the time series.

To verify the model's statistical validity, both stationarity and invertibility conditions were evaluated for the constrained SARIMA (0,1,[1,2,4])(0,1,1)₁₂ model. Regarding stationarity, because the model incorporates zero AR parameters ($p = 0$, $P = 0$), stationarity is inherently satisfied following first-order regular and seasonal differencing ($d = 1$, $D = 1$). Regarding invertibility, the roots of the MA characteristic polynomials were evaluated. The non-seasonal MA polynomial yielded a maximum inverse root modulus of 0.89, while the seasonal MA parameter ($\theta_1 = 0.92$) yielded an inverse root modulus of

0.99. Since all inverse root moduli are strictly less than 1.0, the selected model fully satisfies both stationarity and invertibility constraints.

For the month-wise ARMA model, the parameters for the seven candidate models for each monthly series were estimated utilizing the MLE algorithm. The best model gaining the lowest values for the information criteria (AIC and BIC) was nominated, provided that all estimated parameters (ϕ , θ) were statistically significant at a significance level of $p < 0.05$.

Based on these criteria, the optimal model for each month was nominated, as demonstrated in Table 6 and Figure 5. They illustrate the variation in the candidate models' month-wise ARMA (p,q), parameter values, and AIC and BIC values across the months. The models varied between a month-wise ARMA (1,1) model for January, a month-wise ARMA (2,0) model in March and October, and a month-wise ARMA (1,0) for other months, reflecting the variation in the evaporation time series memory. For each month, the reported model minimized the selected information criterion among the pre-specified candidates while retaining statistically supported parameters.

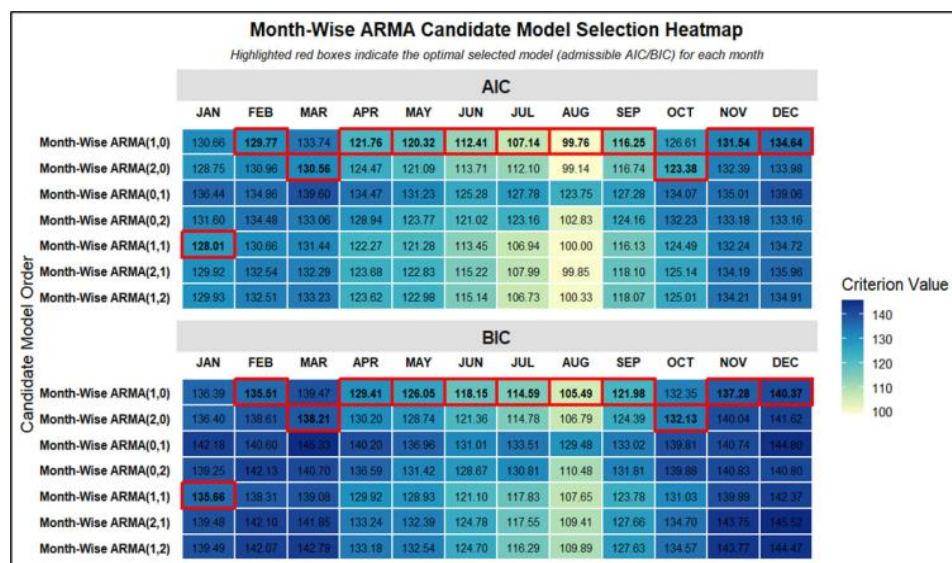


Figure 5. Candidate model selection heatmap for month-wise Autoregressive Moving Average (ARMA) based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) criteria

Note: Highlighted red boxes denote the optimal selected model (minimizing both AIC and BIC) for each monthly series.

Table 6. Estimated parameters and selection criteria for best-fit month-wise Autoregressive Moving Average (ARMA) models of monthly pan evaporation depth (mm)

Month	Month-Wise ARMA model (p,q)	ϕ_1	Sig.	ϕ_2	Sig.	θ_1	Sig.	AIC	BIC
January	(1,1)	0.87	<0.001			0.49	0.01	128.01	135.66
February	(1,0)	0.54	<0.001					129.77	135.51
March	(2,0)	0.35	0.013	0.31	0.028			130.56	138.21
April	(1,0)	0.71	<0.001					121.76	129.41
May	(1,0)	0.65	<0.001					120.32	126.05
June	(1,0)	0.68	<0.001					112.41	118.15
July	(1,0)	0.84	<0.001					107.14	114.59
August	(1,0)	0.79	<0.001					99.76	105.49
September	(1,0)	0.69	<0.001					116.25	121.98
October	(2,0)	0.37	0.003	0.47	0.001			123.38	132.13
November	(1,0)	0.53	<0.001					131.54	137.28
December	(1,0)	0.53	<0.001					134.64	140.37

Note: Each reported entry corresponds to the top-ranked model (Rank 1) selected for that calendar month. The comprehensive relative ranking across all 84 candidate monthly configurations (12 months $\times$ 7 candidate structures) is graphically synthesized in the selection heatmap (Figure 5), where optimal structures minimize both AIC and BIC with parameters significant at $p < 0.05$. Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC).

3.4 Diagnostic checking

After estimating the parameters of the SARIMA constrained model, tests were conducted on the model's residuals to confirm its suitability. The Ljung-Box test demonstrated that the residuals exhibited no residual autocorrelation, yielding $Q(18) = 14.213$ ($df = 14, p = 0.434$), which exceeds the significance level of 0.05. This confirms the acceptance of the null hypothesis, which involves the independence of the residuals and their representation as "white noise". ACF and PACF plots for the residuals were also verified, with all coefficients falling within the confidence interval, thus confirming the model's lack of any unexplained patterns.

The histogram of residuals, demonstrated in Figure 6(a), confirms that the error distribution matches the normal distribution curve. The normal Q-Q plot in Figure 6(b) demonstrates that the experimental points accurately fit the straight line, confirming the conformity of the model residuals to the theoretical normal distribution.

Furthermore, the normality of the residuals hypothesis was tested to ensure the accuracy of the prediction intervals. The Kolmogorov-Smirnov test yielded a probability value of 0.200, while the Shapiro-Wilk test recorded a probability value of 0.095. In addition to the skewness (0.001) and kurtosis (0.635) values presented in Table 7, these results indicate that the residual distribution does not deviate from the normal distribution. This confirms the efficiency and capability of the constrained model in providing predictions for monthly pan evaporation depth in the study area.

Table 7. Residual diagnostics and normality test of Seasonal Autoregressive Integrated Moving Average (SARIMA) model of monthly pan evaporation depth

Kolmogorov-Smirnov		Shapiro-Wilk		Skewness	Kurtosis
Statistic	Sig.	Statistic	Sig.		
0.029	0.200	0.996	0.095	0.001	0.635

For the month-wise ARMA models, diagnostic tests were performed on the residuals for each 12-month model. The Ljung-Box test demonstrated that the significance level for all months exceeded 0.05 (Table 8), confirming the absence of autocorrelation in the residuals and indicating that they represent "white noise", thereby indicating the efficiency of

the nominated models in extracting all available time information.

Furthermore, the distribution of the residuals was verified utilizing the Kolmogorov-Smirnov and Shapiro-Wilk tests. Table 8 demonstrate the results for all months demonstrated probability ($p > 0.05$), confirming that they follow a normal distribution. Figure 7 demonstrates the normal Q-Q plot for the residuals over 12 months. Together, the Ljung-Box and distributional diagnostics indicate that the selected models provide statistically adequate representations of the corresponding subseries within the training sample.

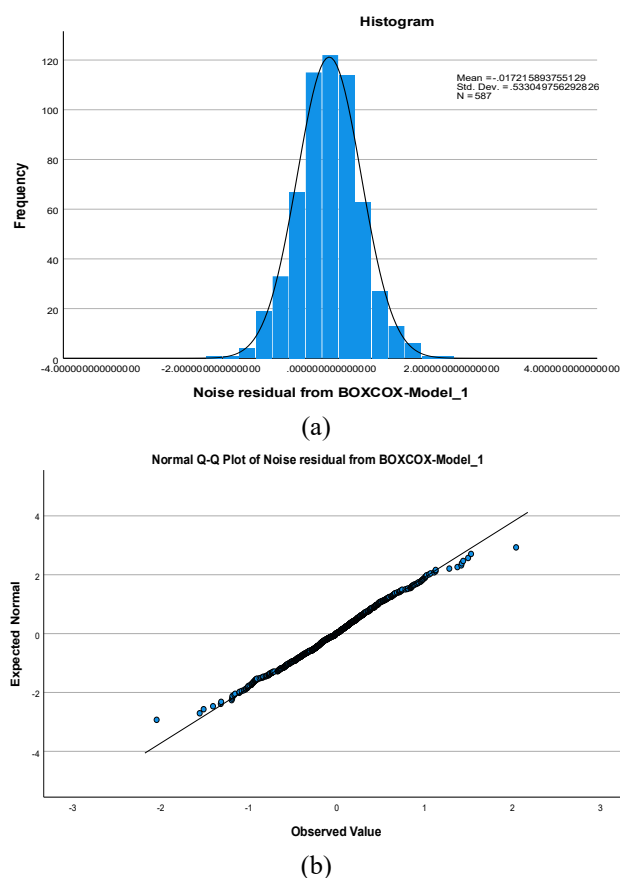
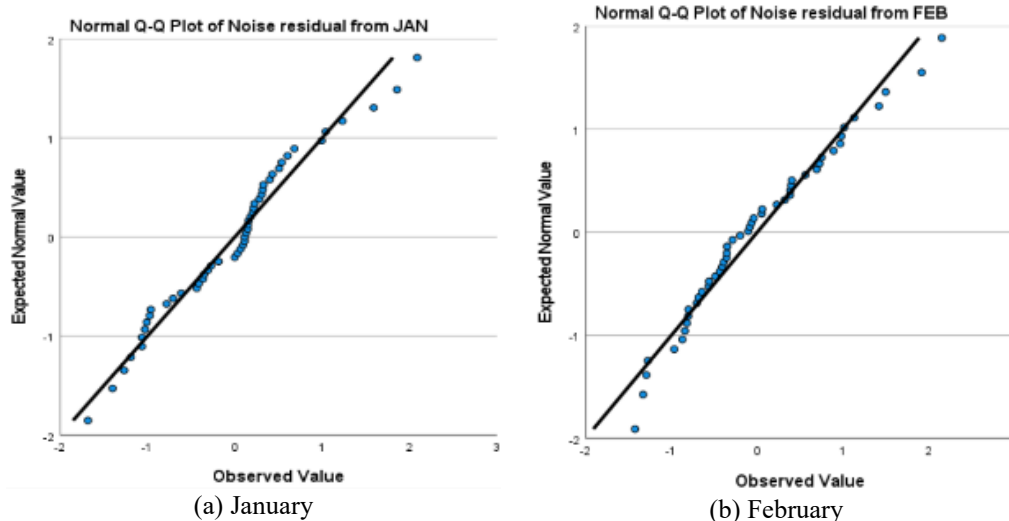
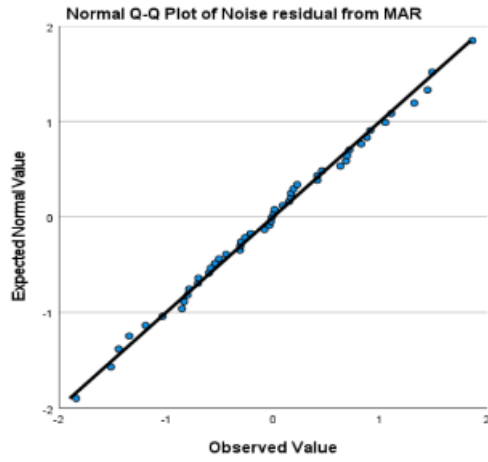
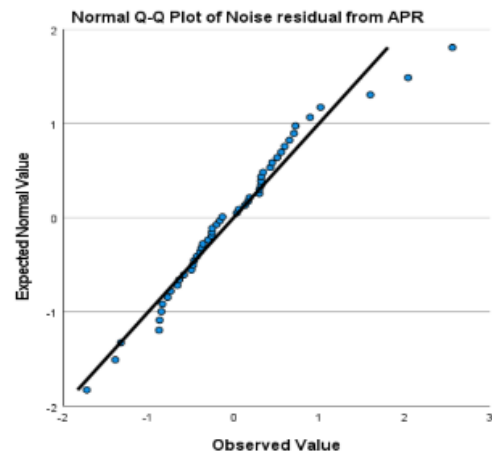


Figure 6. (a) The histogram of residuals ($N = 587$ effective sample size after regular and seasonal differencing, $d = 1, D = 1$); (b) the normal Q-Q plot of residuals for Seasonal Autoregressive Integrated Moving Average (SARIMA)

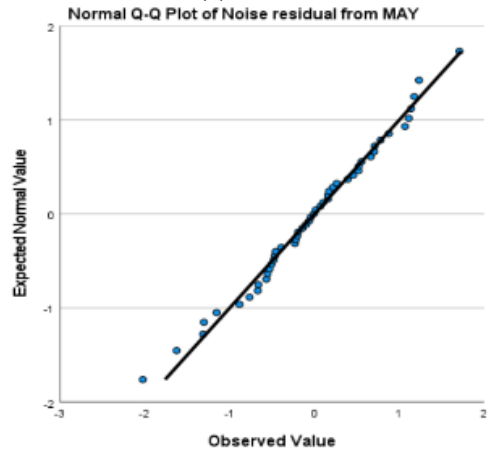




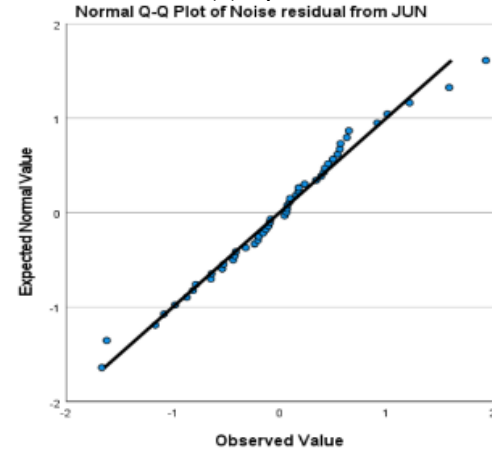
(c) March



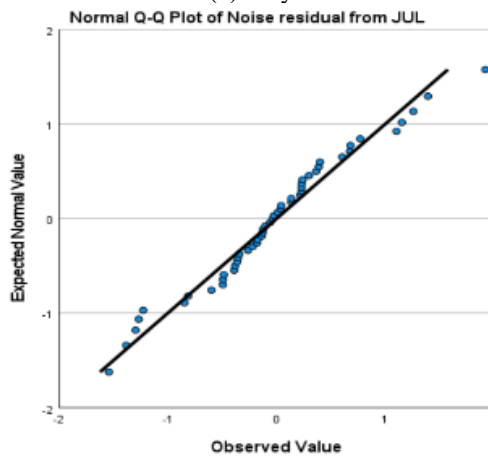
(d) April



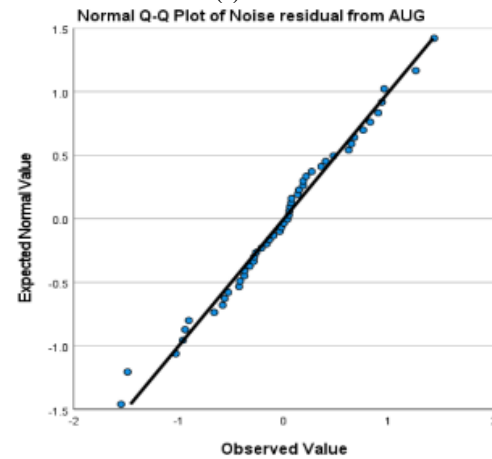
(e) May



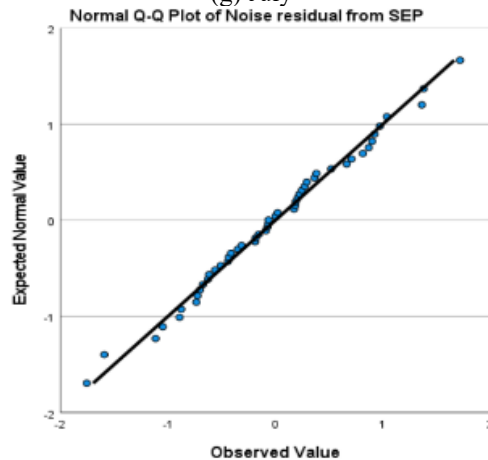
(f) June



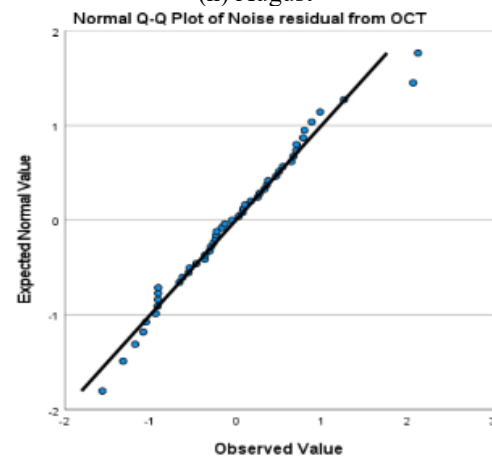
(g) July



(h) August



(i) September



(j) October

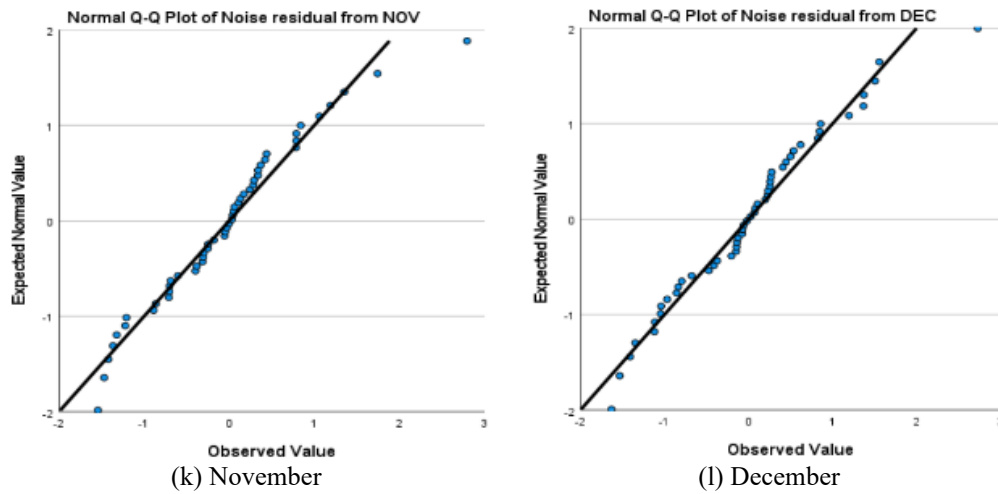


Figure 7. Normal Q-Q plot for residuals of the month-wise Autoregressive Moving Average (ARMA) model over 12 months

Table 8. Diagnostic checking tests for the fitted month-wise Autoregressive Moving Average (ARMA) models of monthly pan evaporation depth

Month	Ljung-Box		Kolmogorov-Smirnova		Shapiro-Wilk	
	Statistics	Sig.	Statistics	Sig.	Statistics	Sig.
January	25.777	0.057	0.110	0.184	0.967	0.180
February	24.485	0.107	0.115	0.093	0.966	0.158
March	7.410	0.965	0.061	0.200	0.993	0.987
April	9.520	0.996	0.090	0.200	0.987	0.857
May	24.684	0.102	0.075	0.200	0.985	0.759
June	10.155	0.897	0.116	0.089	0.968	0.199
July	13.700	0.685	0.094	0.200	0.973	0.292
August	17.012	0.454	0.079	0.200	0.987	0.835
September	20.281	0.260	0.061	0.200	0.991	0.962
October	15.418	0.421	0.083	0.200	0.981	0.612
November	18.149	0.091	0.086	0.200	0.980	0.544
December	13.061	0.732	0.110	0.182	0.974	0.349

Table 9. Comparative performance metrics of Seasonal Autoregressive Integrated Moving Average (SARIMA), month-wise Autoregressive Moving Average (ARMA) and benchmark models of monthly pan evaporation depth (mm) during training and validation periods

Series	Model	RMSE (mm)	MAE (mm)	MAPE	sMAPE	NSE	Pearson's R ²
1968–2017 (Training data)	SARIMA	34.70	25.10	10.6%	10.3%	0.962	96.3%
	Month-wise ARMA	52.50	36.48	15.8%	12.9%	0.910	92.3%
2018–2024 (Validation data)	SARIMA	62.40	47.80	27.4%	24.0%	0.671	86.2%
	Month-wise ARMA	62.21	48.46	30.5%	25.7%	0.671	85.4%
	Seasonal naive benchmark	65.86	50.15	31.1%	26.0%	0.610	85.5%

Note: RMSE = root mean square error; MAE = mean absolute error; MAPE = mean absolute percentage error; sMAPE = symmetric mean absolute percentage error; NSE = Nash-Sutcliffe efficiency.

3.5 Model performance comparison

During the training period (1968–2017), both models demonstrated high fitting performance, with the SARIMA model recording lower error metrics (RMSE = 34.70 mm, MAE = 25.10 mm, sMAPE = 10.3%), a higher NSE of 0.962 and a higher Pearson's R² of 96.3%.

The predictive performance of the SARIMA and month-wise ARMA models was evaluated during the validation period (2018–2024) utilizing several statistical metrics (Table 9).

Overall, both approaches demonstrated broadly comparable predictive efficiency. The month-wise ARMA model achieved a marginally lower RMSE (62.21 mm vs. 62.40 mm), whereas SARIMA produced lower MAE (47.80 mm vs. 48.46 mm) and MAPE (27.4% vs. 30.5%), and lower sMAPE (24.0% vs. 25.7%). Both models exhibited strong hydrological efficiency

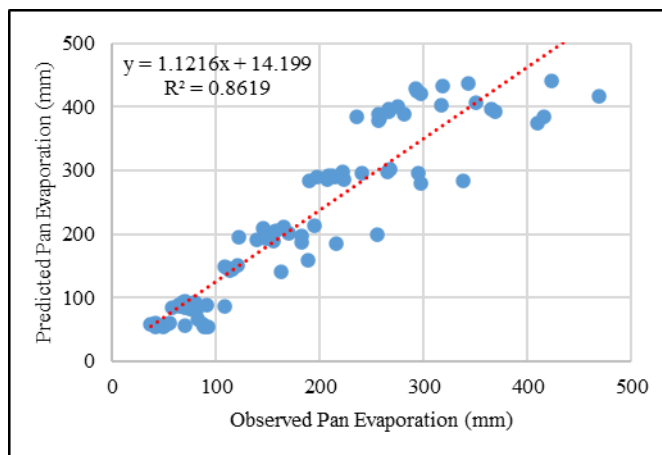
with NSE of 0.671 alongside high structural alignment with observed data, as indicated by R² = 86.2% for SARIMA and 85.4% for month-wise ARMA. This close agreement is clearly illustrated in the observed-versus-predicted scatter plots (Figures 8(a) and (b)), showing tight clustering along the linear regression lines across low-to-moderate evaporation depths, with constrained dispersion during peak summer months.

To evaluate the practical skill of the calibrated stochastic models, a seasonal naive baseline was evaluated across the 2018–2024 validation period (Table 9). The seasonal naive model yielded higher error metrics (RMSE = 65.86 mm, MAPE = 31.1%) and lower hydrological skill (NSE = 0.610). While its high correlation (R² = 85.5%) simply reflects the deterministic intra-annual seasonality driven by radiative and thermal variations, both SARIMA and month-wise ARMA improved the NSE by +0.061 and reduced RMSE by approximately 3.5 mm. This demonstrates that the proposed

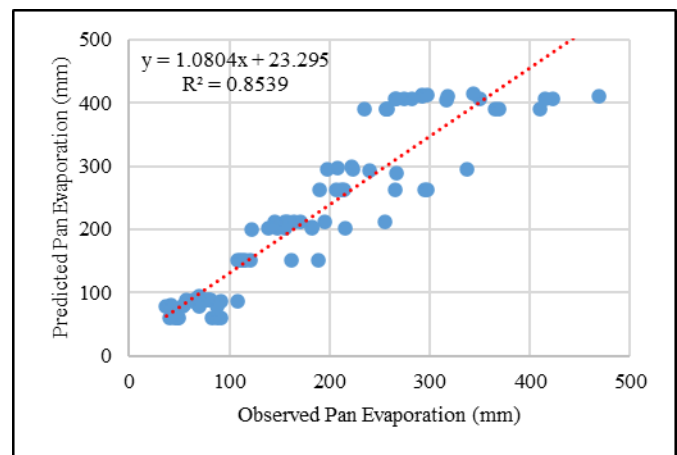
stochastic formulations provide tangible added value by capturing inter-annual memory beyond simple seasonal recurrence.

To formally evaluate whether the observed differences in predictive accuracy are statistically significant, the DM test was conducted on the validation forecast errors. The test yielded $DM = 0.0221$, $p = 0.9824$ under squared-error loss and

$DM = -0.4331$, $p = 0.6661$ under absolute error loss, confirming that the difference in forecast accuracy between the two models is statistically insignificant ($p > 0.05$) [49]. Therefore, SARIMA was retained as the reference model for long-horizon forecasting due to its parsimony, structural simplicity and lower percentage error metrics.



(a)



(b)

Figure 8. Scatter plots of observed versus predicted monthly pan evaporation depth (mm) during the 2018–2024 validation period: (a) constrained Seasonal Autoregressive Integrated Moving Average (SARIMA) $(0,1,[1,2,4])(0,1,1)_{12}$ model; (b) month-wise Autoregressive Moving Average (ARMA) model

Note: The red dotted lines denote the linear regression fits.

3.6 Forecasting and discussion

It is seen by comparing SARIMA and month-wise ARMA models that these models vary in the way of data representation. As can be seen in Figure 9, the SARIMA model is more similar to real values, as the projections are similar to the actual evaporation data. The month-wise ARMA model (Figure 10) dramatically overestimates the peak evaporation in summer for several years. Although the month-wise ARMA achieved a marginally lower RMSE (62.21 mm vs. 62.40 mm), this summer over-forecasting accounts for its higher absolute and percentage error metrics (MAE = 48.46 mm vs. 47.80 mm, MAPE = 30.5% vs. 27.4%), indicating that independent cyclic subseries modeling is more sensitive to interannual peak anomalies.

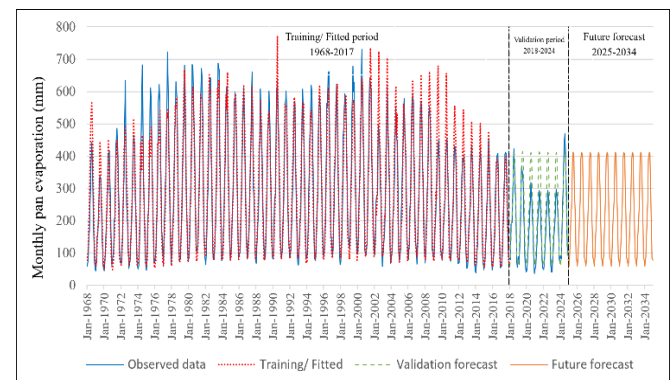


Figure 10. Observed, fitted, and forecasted monthly pan evaporation depth (mm) using the month-wise Autoregressive Moving Average (ARMA) model across training (1968–2017), validation (2018–2024), and future forecast (2025–2034)

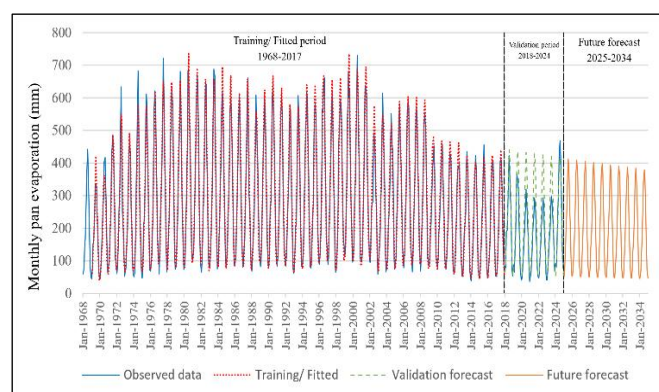


Figure 9. Observed, fitted, and forecasted monthly pan evaporation depth (mm) using Seasonal Autoregressive Integrated Moving Average (SARIMA) model across training (1968–2017), validation (2018–2024), and future forecast (2025–2034) periods

The projected plots over the next 10 years (January 2025–December 2034) demonstrate a steady seasonal pattern with obvious oscillations after yearly temperature changes. The two models predicted the continuity of this seasonal pattern. The SARIMA model provides a more structurally balanced representation across seasonal transitions. Therefore, given that the forecast accuracy differences are statistically insignificant ($DM p > 0.05$), SARIMA serves as a pragmatic and parsimonious baseline reference for regional evaporation risk assessment. The convergence in the statistical performance indicates that the evaporation data series in Najaf has high autocorrelation features and the Box-Jenkins approach is the most efficient framework for processing this sort of data.

The forecasts retain the estimated seasonal recurrence embedded in the fitted SARIMA structure. This should be interpreted as a model-based continuation of historical seasonal dependence, not as evidence that the underlying climate process is mean-reverting.

The reduction in the predictive accuracy from calibration to validation indicates possible temporal non-stationarity or structural change that is not captured by the univariate models. Because the present study does not explicitly model meteorological covariates or perform causal attribution, the source of this degradation cannot be identified from the current analysis. Future work should test structural breaks and incorporate exogenous climatic predictors.

95% prediction intervals were computed for each monthly forecast on the original evaporation scale. Figure 11 presents the validation (2018–2024) and future forecast periods separately, with the future prediction interval shown for January 2025–December 2034. The forecast indicates that evaporative losses remain concentrated in July, with a projected median monthly peak of 396 mm. Over the 10-year decadal horizon, the 95% prediction interval spans 185.5–746 mm at mid-horizon in 2029 and progressively widens to approximately 850 mm by 2034 as multi-step forecast uncertainty accumulates. Importantly, these decadal projections rely on the fundamental assumption that historical stochastic dependence structures and climatic variance remain approximately stationary. In the context of accelerating climate warming and regional drought trends across Iraq, this assumption represents a primary structural limitation, as non-stationary thermal shifts could drive actual future evaporative peaks toward or beyond the upper prediction limits. Consequently, these projections serve as a probabilistic baseline under historical stationarity rather than deterministic climate-change projections. Because 1 mm of evaporation over 1 km² corresponds to approximately 1,000 m³ of water loss, this peak corresponds to a volumetric loss of approximately 396,000 m³/km², allowing forecasts to be converted directly to volumetric exposure for reservoirs or open-water surfaces of known area.

56 guidelines [51] can be applied to convert forecasted monthly pan evaporation into direct open-water evaporation losses. This allows regional authorities, such as the Directorate of Water Resources in Al-Najaf, to quantify gross surface storage depletion in local water bodies, including the Bahr Al-Najaf depression and storage basins, and dynamically adjust summer water-release rules to offset anticipated volumetric deficits. In irrigation management, projected monthly evaporative rates provide essential input boundaries for estimating crop water requirements (ET_c) and determining leaching fractions for the surrounding agrarian zones, thereby optimizing irrigation scheduling and safeguarding root-zone soils from hyper-evaporative capillary salinization during peak summer months [9]. Furthermore, regarding drought monitoring, the upper bound of the 95% prediction interval (reaching up to 746 mm during extreme July conditions) serves as a critical probabilistic benchmark for evaporative stress testing. Tracking actual observations against these probabilistic bounds provides regional planners with an early-warning mechanism to trigger pre-emptive drought contingency measures before extreme evaporative deficits exhaust local water reserves [4].

4. CONCLUSIONS

The research reveals that the evaporation time series for the Najaf area have a significant seasonal autocorrelation, making them naturally suitable for stochastic characterization utilizing the Box-Jenkins approach. During the 2018–2024 validation period, SARIMA and month-wise ARMA models produced comparable forecast errors, although the ranking depended on the metric: the month-specific model had a marginally lower RMSE (62.21 mm vs. 62.40 mm), while SARIMA achieved lower MAE, MAPE and sMAPE. Formal DM testing confirmed that this difference in forecast accuracy between the two models is statistically insignificant ($p > 0.05$).

A key feature of the month-wise ARMA model is that each calendar month is assigned its own interannual dependence structure, allowing heterogeneous month-specific dynamics. Its limitation is that it does not model dependence between adjacent months. SARIMA, in contrast, provides a single parsimonious seasonal structure and yielded lower MAE, MAPE and sMAPE in validation, motivating its use for the long-horizon reference forecast.

The long-horizon baseline forecast preserves pronounced seasonal peaks and troughs, while uncertainty widens with the forecast horizon. Crucially, these projections should be interpreted as a stochastic planning reference for atmospheric evaporation demand variability rather than a direct or deterministic assessment of future water losses, as actual volumetric losses remain contingent upon open-water surface dynamics and reservoir operational regimes. SARIMA and the month-specific model provide complementary statistical perspectives on temporal persistence and seasonal heterogeneity, offering regional water authorities a probabilistic baseline for water-resource risk assessment.

Despite the structured statistical workflow, the study has several limitations. First, it relies on observations from a single meteorological station in Najaf, limiting spatial generalizability. Second, the univariate models omit exogenous drivers such as temperature, wind speed, relative humidity, and radiation; these could be incorporated through SARIMAX or dynamic-regression models. Third, while the

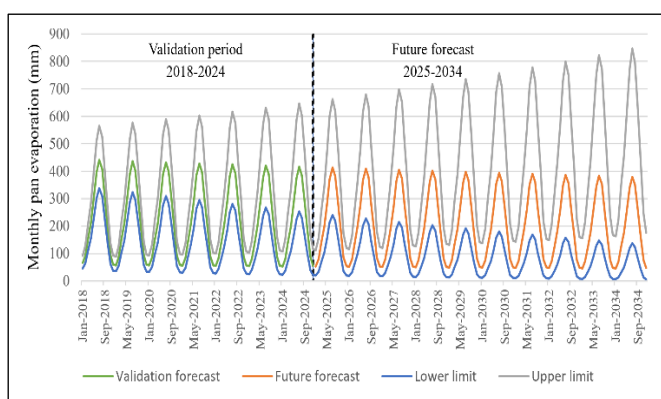


Figure 11. Monthly pan evaporation depth (mm) forecasts using Seasonal Autoregressive Integrated Moving Average (SARIMA) with 95% prediction intervals for the validation period (2018–2024) and 10-year future horizon (2025–2034)

To translate these stochastic projections into operational water-resource planning, the generated forecasts establish actionable pathways across three primary management domains. In reservoir water allocation, regional pan coefficients ($K_p = 0.70\text{--}0.75$) established under standard FAO-

calibrated stochastic models offer structural transparency, they inherently assume linear temporal dependence, omitting complex nonlinear climatic interactions that data-driven machine-learning algorithms (e.g., LSTM and XGBoost) can capture. Future work should systematically benchmark these linear stochastic baselines against nonlinear machine-learning architectures under the identical time-aware protocol.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the Iraqi Meteorological Organization and Seismology (IMOS), Ministry of Transport, for providing the historical pan evaporation records utilized in this study.

REFERENCES

- [1] Aljuhaishy, S.S.D., Naimi, S., Awchi, T.A., Manfreda, S. (2025). Impact of climate change on evaporation in Mosul Dam reservoir-Iraq. *AIP Conference Proceedings*, 3303(1): 050005. <https://doi.org/10.1063/5.0264494>
- [2] Al-Hasani, B., Abdellatif, M., Carnacina, I., Harris, C., Al-Quraishi, A.M.F., Al-Shammari, M.M.A. (2025). Forecasting evaporation trends amid climate change for sustainable water management in semi-arid regions. *Water*, 17(7): 1039. <https://doi.org/10.3390/w17071039>
- [3] Shamkhi, M.S., Al-Badry, H.J. (2022). Soil texture distribution for east Wasit province, Iraq. *IOP Conference Series: Earth and Environmental Science*, 961(1): 012073. <https://doi.org/10.1088/1755-1315/961/1/012073>
- [4] Vicente-Serrano, S.M., McVicar, T.R., Miralles, D.G., Yang, Y., Tomas-Burguera, M. (2020). Unraveling the influence of atmospheric evaporative demand on drought and its response to climate change. *WIREs Climate Change*, 11(2): e632. <https://doi.org/10.1002/wcc.632>
- [5] Farhan, A.A., Abed, B.S. (2021). Estimation of surface runoff to Bahr Al-Najaf. *Journal of Engineering*, 27(9): 51-63. <https://doi.org/10.31026/j.eng.2021.09.05>
- [6] Mahdi, H.H., Musa, T.A., Al-Rammahi, Z.A.A., Mahmood, E.J. (2022). Forecasting the future drought indices due to the effects of climate change in Al Najaf City, Iraq. *IOP Conference Series: Earth and Environmental Science*, 961(1): 012040. <https://doi.org/10.1088/1755-1315/961/1/012040>
- [7] Al-Ansari, N.A. (2013). Management of water resources in Iraq: Perspectives and prognoses. *Engineering*, 5(8): 667-684. <https://doi.org/10.4236/eng.2013.58080>
- [8] Allawi, M.F., Abdulhameed, U.H., Sahab, M.F., Sulaiman, S.O. (2026). Estimation of evaporation rate using advanced methods. *An-Najah University Journal for Research-A (Natural Sciences)*, 40(1): 65-70. <https://doi.org/10.35552/anujr.a.40.1.2441>
- [9] Phesa, M., Woyessa, Y., Mbatha, N. (2022). Evapotranspiration prediction using ARIMA, ANN and hybrid models for optimum water use in agriculture: A case study of Keiskammahoek Irrigation Scheme, Eastern Cape, South Africa. In *Proceedings of the 3rd International Symposium on Water, Ecology and Environment ISWEE*, Beijing, China, pp. 276-286. <https://doi.org/10.5220/0012009000003536>
- [10] Shimi, M., Najjarchi, M., Khalili, K., Hezavei, E., Mirhoseyni, S.M. (2020). Investigation of the accuracy of linear and nonlinear time series models in modeling and forecasting of pan evaporation in Iran. *Arabian Journal of Geosciences*, 13: 59. <https://doi.org/10.1007/s12517-019-5031-7>
- [11] Machiwal, D., Jha, M.K. (2012). *Hydrologic Time Series Analysis: Theory and Practice*. Springer Dordrecht. <https://doi.org/10.1007/978-94-007-1861-6>
- [12] Shad, M., Sharma, Y.D., Singh, A. (2022). Forecasting of monthly relative humidity in Delhi, India, using SARIMA and ANN models. *Modeling Earth Systems and Environment*, 8: 4843-4851. <https://doi.org/10.1007/s40808-022-01385-8>
- [13] Al-Ataby, I.K., Hassan, B.J., Al-Timimi, Y.K. (2025). Time series decomposition and prediction of the monthly temperature of some stations in Iraq. *Iraqi Geological Journal*, 58(1C): 180-187. <https://doi.org/10.46717/igj.58.1C.14ms-2025-3-29>
- [14] Box, G.E.P., Jenkins, G.M. (1968). Some recent advances in forecasting and control. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 17(2): 91-109. <https://doi.org/10.2307/2985674>
- [15] Al-Mukhtar, M. (2021). Modeling the monthly pan evaporation rates using artificial intelligence methods: A case study in Iraq. *Environmental Earth Sciences*, 80: 39. <https://doi.org/10.1007/s12665-020-09337-0>
- [16] Landaras, G., Ortiz-Barredo, A., López, J.J. (2009). Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. *Journal of Irrigation and Drainage Engineering*, 135(3): 323-334. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000008](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000008)
- [17] Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). John Wiley & Sons.
- [18] Hipel, K.W., McLeod, A.I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier.
- [19] Sule, B.F., Ajala, O.F. (2017). Long-term estimates of reservoir evaporation using ARIMA model and impact on water supply: A case study of Erinle dam, Osun State, Nigeria. *International Journal of Sciences*, 6(9): 29-38. <https://doi.org/10.18483/ijSci.1331>
- [20] Benchaiba, L., Houichi, L., Amarchi, H. (2022). Stochastic alternatives to models based on the Dalton's approach for estimating monthly pan evaporation from reservoir dams in Algeria. *LARHYSS Journal*, 52: 49-59. <https://www.larhyss.net/ojs/index.php/larhyss/article/view/858/860>
- [21] Mamudu, L., Yahaya, A., Danazumi, S., Mamudu, L. (2021). Application of seasonal autoregressive integrated moving average (SARIMA) for flows of River Kaduna. *Nigerian Journal of Engineering*, 28(2): 84-90. <https://njeabu.com.ng/?mno=108688>
- [22] Dabral, P.P., Tabing, I. (2020). Modelling and forecasting of monthly rainfall and temperature time series using SARIMA for trend detection-a case study of Umiam, Meghalaya (India). *International Journal of Environment and Climate Change*, 10(11): 155-172. <https://doi.org/10.9734/IJECC/2020/v10i1130276>
- [23] Rezaei, R., Shabri, A. (2023). Using the ARIMA/SARIMA model for Afghanistan's drought forecasting based on standardized precipitation index. *Matematika*, 39(3): 239-261.

- <https://doi.org/10.11113/matematika.v39.n3.1478>
- [24] Wang, H.R., Wang, C., Lin, X., Kang, J. (2014). An improved ARIMA model for precipitation simulations. *Nonlinear Processes in Geophysics*, 21(6): 1159-1168. <https://doi.org/10.5194/npg-21-1159-2014>
- [25] Albazzaz, Z.M., Fadhil, N.S. (2024). Using hybrid regression tree and ARIMA model with wavelet transforms for evaporation time series forecasting. *Iraqi Journal of Statistical Sciences*, 21(2): 125-134. <https://doi.org/10.33899/ijoss.2024.185246>
- [26] Başakın, E.E., Ekmekcioğlu, Ö., Özger, M. (2023). Providing a comprehensive understanding of missing data imputation processes in evapotranspiration-related research: A systematic literature review. *Hydrological Sciences Journal*, 68(14): 2089-2104. <https://doi.org/10.1080/02626667.2023.2249456>
- [27] Gao, Y.B., Merz, C., Lischeid, G., Schneider, M. (2018). A review on missing hydrological data processing. *Environmental Earth Sciences*, 77: 47. <https://doi.org/10.1007/s12665-018-7228-6>
- [28] World Meteorological Organization. (2008). Guide to Instruments and Methods of Observation (WMO-No. 8). World Meteorological Organization, Geneva, Switzerland.
- [29] Alwan, I.A., Ziboon, A.R.T., Khalaf, A.G. (2019). Utilization of reconnaissance drought index (RDI) for monitoring of meteorological drought over middle Euphrates region during the period from 1988 to 2017. *IOP Conference Series: Materials Science and Engineering*, 518(2): 022035. <https://doi.org/10.1088/1757-899X/518/2/022035>
- [30] Younis, A.M., Jasim, H.M. (2024). Regional drought assessment based on the standard precipitation index (SPI) and precipitation concentration index (PCI) in middle part of Iraq by using GIS. *IOP Conference Series: Earth and Environmental Science*, 1371(8): 082021. <https://doi.org/10.1088/1755-1315/1371/8/082021>
- [31] Kadhum, Z., Alhameedawi, A., Hamoodi, M. (2025). Prediction of reconnaissance drought index and standardized precipitation index for drought analysis. *Engineering and Technology Journal*, 43(11): 905-919. <https://doi.org/10.30684/etj.2025.159196.1943>
- [32] Montgomery, D.C., Jennings, C.L., Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons.
- [33] Messaoudi, A., Attou, F., Ghennam, K., Nabi, F., Mostefaoui, H. (2026). Time series analysis and modeling of physicochemical water parameters in Ghrib dam, Medea, Algeria. *Ecological Engineering & Environmental Technology*, 27(1): 314-324. <https://doi.org/10.12912/27197050/215465>
- [34] Box, G.E.P., Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2): 211-243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- [35] Saleh, M.S., Hamad, H.T. (2020). Forecasting the annual flow rate of the Tigris River using stochastic modelling. *Periodicals of Engineering and Natural Sciences*, 8(4): 2035-2044. <https://doi.org/10.21533/pen.v8.i4.1371>
- [36] Shumway, R.H., Stoffer, D.S. (2017). *Time Series Analysis and its Applications: With R Examples* (4th ed.). Springer, Cham. <https://doi.org/10.1007/978-3-319-52452-8>
- [37] Rezaei, R., Shabri, A. (2023). Drought forecasting using W-ARIMA model with standardized precipitation index. *Journal of Water and Climate Change*, 14(9): 3345-3367. <https://doi.org/10.2166/wcc.2023.431>
- [38] Dabral, P.P., Murry, M.Z. (2017). Modelling and forecasting of rainfall time series using SARIMA. *Environmental Processes*, 4: 399-419. <https://doi.org/10.1007/s40710-017-0226-y>
- [39] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press. <https://doi.org/10.1515/9780691218632>
- [40] Nugroho, A., Rahayu, D.E., Sutiarso, L., Murtiningrum, Fallah, M.A.F., Okayasu, T. (2021). Development of short-term evapotranspiration forecasting model using time series method for supporting the precision agriculture management in tropics. *IOP Conference Series: Earth and Environmental Science*, 653(1): 012034. <https://doi.org/10.1088/1755-1315/653/1/012034>
- [41] Ali, S. (2013). Time series analysis of Baghdad rainfall using ARIMA method. *Iraqi Journal of Science*, 54(4): 1136-1142. <https://ijs.uobaghdad.edu.iq/index.php/eijs/article/view/12433>
- [42] Machiwal, D., Jha, M.K. (2008). Comparative evaluation of statistical tests for time series analysis: Application to hydrological time series / Evaluation comparative de tests statistiques pour l'analyse de séries temporelles: Application à des séries temporelles hydrologiques. *Hydrological Sciences Journal*, 53(2): 353-366. <https://doi.org/10.1623/hysj.53.2.353>
- [43] Chai, T., Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3): 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- [44] Willmott, C.J., Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30: 79-82. <https://doi.org/10.3354/cr030079>
- [45] Makridakis, S. (1993). Accuracy measures: Theoretical and practical concerns. *International Journal of Forecasting*, 9(4): 527-529. [https://doi.org/10.1016/0169-2070\(93\)90079-3](https://doi.org/10.1016/0169-2070(93)90079-3)
- [46] Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3): 885-900. <https://doi.org/10.13031/2013.23153>
- [47] Legates, D.R., McCabe, G.J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1): 233-241. <https://doi.org/10.1029/1998WR900018>
- [48] Lailiyah, I., Wardhani, M.K., Cahya, I., Kurniawan, F., Febriansyah, R. (2026). SARIMA model: Time series forecasting of temperature and salinity in Madura Strait up to 2030. *IOP Conference Series: Earth and Environmental Science*, 1585(1): 012013. <https://doi.org/10.1088/1755-1315/1585/1/012013>
- [49] Diebold, F.X., Mariano, R.S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1): 134-144. <https://doi.org/10.1198/073500102753410444>

- [50] Hyndman, R.J., Athanasopoulos, G. (2021). Forecasting: Principles and Practice. OTexts, Melbourne, Australia. <https://otexts.com/fpp3/>.
- [51] Allen, R.G., Pereira, L.S., Raes, D., Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and Drainage Paper 56. FAO, ROME, 300(9): D05109. https://www.avwatermaster.org/filingdocs/195/70653/172618e_5xAGWAx8.pdf.

NOMENCLATURE

ACF	autocorrelation function
ADF	augmented Dickey-Fuller test
AIC	akaike information criterion (dimensionless)
AR	autoregressive component/order family of a time-series model (dimensionless)
ARIMA	autoregressive integrated moving average model
ARMA	autoregressive moving average model
B	backward (backshift) operator (dimensionless)
BIC	Bayesian information criterion (dimensionless)
D	order of seasonal differencing in SARIMA (dimensionless)
d	order of regular (non-seasonal) differencing in SARIMA (dimensionless)
df	degrees of freedom (dimensionless)
DM	Diebold-Mariano test statistic (dimensionless)
d_t	Loss differential at time t (mm^2 for squared error, mm for absolute error)
$\bar{d}$	Mean loss differential across the validation period
$e_{1,t}$	forecast error of model 1 at observation/time t (mm)
$e_{2,t}$	forecast error of model 2 at observation/time t (mm)
ET_C	crop water requirement under standard conditions (mm month^{-1})
h	lag (dimensionless)
i	observation index (dimensionless)
IMOS	Iraqi Meteorological Organization and Seismology
j	lag index (dimensionless)
k	number of estimated model parameters (dimensionless)
k_p	pan evaporation coefficient to estimate open-water loss (dimensionless)
L	maximum-likelihood function/value (dimensionless)
m	calendar-month index (dimensionless)
MA	moving-average (dimensionless)
MAE	mean absolute error (mm)
MAPE	mean absolute percentage error (%)
MLE	maximum likelihood estimation
n, N	number of observations (dimensionless)
NSE	Nash-Sutcliffe efficiency (dimensionless)

P	order of the seasonal autoregressive component (dimensionless)
PACF	partial autocorrelation function
Q	order of the seasonal moving-average (dimensionless)
R^2	Pearson's coefficient of determination (dimensionless)
RMSE	root mean square error (mm)
S	length of the seasonal cycle (dimensionless)
SARIMA	seasonal autoregressive integrated moving average model
sMAPE	Symmetric mean absolute percentage error (%)
t	time (monthly)
WMO	World Meteorological Organization
X_t	original observed monthly pan evaporation at time t (mm)
$X_{\gamma, m}^*$	transformed evaporation value for year y and month m (mm)
Y_i	observed monthly pan evaporation value at observation i (mm)
Y_{i-12}	observed monthly pan evaporation value 12 months before observation i (mm)
Y_t	observed monthly pan evaporation depth at time t (mm)
Y^{λ}	Box-Cox-transformed value of the observed series (dimensionless)
$\bar{Y}$	mean observed monthly pan evaporation value (mm)
$\hat{Y}_i$	predicted pan evaporation value corresponding to observation i (mm)
$\hat{Y}_t$	predicted pan evaporation value at time t (mm)
$Z_{m, \gamma}$	standardized evaporation value for calendar month in year (dimensionless)
Z_{m, γ_j}	lagged standardized evaporation value for calendar month, year (dimensionless)

Greek symbols

ε_t	white-noise error at time t
$\varepsilon_{m, \gamma}$	white-noise error for calendar month m in year y
μ_m	mean pan evaporation for calendar month m (mm)
$\mu_{\hat{Y}}$	mean of predicted pan evaporation value (mm)
Φ_j	seasonal autoregressive coefficient
$\Phi_p(B^s)$	seasonal autoregressive polynomial
ϕ_j	non-seasonal autoregressive coefficient
$\phi_p(B)$	non-seasonal autoregressive polynomial
σ_m	standard deviation of evaporation for calendar month m (mm)
λ	optimal Box-Cox power transformation parameter
Θ_j	seasonal moving-average coefficient at seasonal lag j
$\Theta_Q(B^s)$	seasonal moving-average polynomial