Water and Soil Management and Modeling

Online ISSN: 2783 - 2546

University of Mohaghegh Ardabili

Development of multiple linear regression models for annual reference evapotranspiration estimation under limited data conditions

Hedieh Ahmadpari[1], Vitaly Khaustov[2], Ata Amini[3*]

[1] Ph.D. candidate, Hydrology of Land, Water Resources, Hydrochemistry, Russian State Hydrometeorological University, Saint Petersburg, Russia.
[2] Candidate of Technical Sciences, Associate Professor at the Department of Engineering Hydrology of the RSHU, Saint Petersburg, Russia.
[3*] Professor, Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj, Iran.

Abstract

Accurate estimation of reference evapotranspiration (ET_0) is essential for agricultural water management, particularly in regions with limited data availability. This study aimed to evaluate multiple linear regression (MLR) models to estimate ET_0 at the annual scale. Meteorological data from the Kuhdasht synoptic station, Iran, for 25 years (1998–2022) were used. ET_0 was calculated using the FAO-56 Penman-Monteith method implemented through the CROPWAT 8.0 software. A total of 31 MLR models were developed using the Regression option from the Analysis ToolPak of Microsoft Excel 2019 to quantify the relationship between ET_0 and climatic variables. Seven statistical indices were used to evaluate the performance of the MLR models in estimating ET_0. Results showed that 16 models achieved very high accuracy, with coefficients of determination (R^2) greater than 0.92. Among single-variable models, wind speed (MLR4) emerged as the strongest predictor, explaining up to 92% of ET_0 variability. Several two-variable models achieved R^2 = 0.92–0.96, and most three-variable models reached R^2 = 0.93–0.97. Four-variable models also performed strongly ($R^2 \approx$ 0.95–0.97), while the five-variable model yielded $R^2 \approx 0.97$, similar to simpler models. Wind speed emerged as the most influential factor, highlighting that well-chosen two- or three-variable models can estimate ET_0 as effectively as more complex alternatives.

Keywords: Reference evapotranspiration, FAO-56 Penman–Monteith, Multiple Linear Regression, Wind speed

Article Type: Research Article
Academic Editor: Raoof Mostafazadeh

*Corresponding Author, E-mail: ata_amini@yahoo.com

Citation: Ahmadpari, H., Khaustov, V., Amini, A. (2026). Development of multiple linear regression models for annual reference evapotranspiration estimation under limited data conditions. Water and Soil Management and Modelling, 6(2) (Special Issue: New Approaches to Water and Soil Management and Modeling), 209-231.
doi: 10.22098/mmws.2025.18810.1726
Received: 11 November 2025, Received in revised form: 30 November 2025, Accepted: 13 December 2025, Published online: 03 June 2026
Water and Soil Management and Modeling, Year 2026, , Vol. 6, No.2 (Special Issue), pp. 209-231
Publisher: University of Mohaghegh Ardabili

1. Introduction

Reference evapotranspiration (ET_0) is a fundamental component in hydrological studies and agricultural water management, because it represents the climatic demand for water by a reference surface and is critical in designing irrigation schedules and managing water resources (Liu et al., 2024). The methods of estimating the evapotranspiration of plants are divided into two main groups: direct and indirect (computational). Various methods, including lysimetric methods, are proposed in the form of direct methods for measuring evapotranspiration, but the use of a lysimeter is not feasible due to the lack of affordable and time-consuming measurement of data in all regions (Ahmadpari et al., 2019a). For this reason, researchers have tried to use indirect methods of estimating evapotranspiration from evaporation pan values or some meteorological data. In all indirect methods that are used to determine the amount of evapotranspiration, the ET_0 value is estimated and, using this, the water requirement of the desired plant is calculated (Ahmadpari et al., 2019). Several methods have been proposed for estimating Reference evapotranspiration; each of them has certain limitations and can be recommended in special conditions. All of these methods are a combination of theoretical concepts and empirical results (Ahmadpari et al., 2017). The FAO-56 Penman–Monteith method is widely accepted as a global standard for estimating Reference evapotranspiration, but it requires comprehensive meteorological data, such as solar radiation, wind speed, air temperature, and humidity, which may not be available in many regions (Khadempour et al., 2017).

In many data-scarce or remote environments, meteorological stations either lack certain sensors or have incomplete records, making it difficult to apply methods like FAO-56 Penman–Monteith directly (Amini & Hesami, 2017). As a general rule, the level of prediction accuracy generally improves with the increase in input parameters (Djaman et al., 2017). The cost of the necessary sensors, the equipment, and the maintenance of meteorological stations, along with the difficult environmental conditions, results in data unavailability or at least scarcity, especially for the developing countries. Therefore, an effort to reduce the quantity of data required for predictive models is highly recommended, provided that satisfactory accuracy is secured (Tegos et al., 2015). To cope with these constraints, multiple linear regression has been explored as a viable alternative because it can use a subset of available variables and yet produce acceptable accuracy (Dimitriadou and Nikolakopoulos, 2022).

Several studies have applied multiple linear regression (MLR) techniques to estimate ET_0 using meteorological variables such as temperature, relative humidity, wind speed, and solar radiation in different climatic regions. In the Brazilian Amazon, site-specific MLR equations using temperature, wind speed, and insolation reproduced FAO-56 Penman–Monteith estimates with good accuracy, especially when three predictors were included (da Silva et al., 2016). Research in Turkey confirmed the strong predictive capacity of multiple linear regression, where models using only air temperature and relative humidity achieved coefficients of determination close to 0.99 (Usta and Gencoglan, 2019). Applications in East Africa also confirmed the usefulness of multiple linear regression. For example, in Ethiopia, regression models based on temperature, wind speed, sunshine duration, and relative humidity closely matched Penman–Monteith reference evapotranspiration, making them suitable for use in data-scarce basins (Yirga, 2019). In Greece, the use of combinations of sunshine duration, mean temperature, solar radiation, net radiation, wind speed, and vapour pressure deficit resulted in an adjusted R^2 of 0.98 and an RMSE of just 0.28 mm day^{-1}, confirming the high accuracy of well-structured regression formulations (Dimitriadou and Nikolakopoulos, 2022). More recently, the potential of MLR has been demonstrated for situations with incomplete climate data, where appropriate substitutions, such as replacing dew point with minimum temperature or applying regional averages for wind speed and humidity, still resulted in acceptable estimates (Koç and Can, 2023). Overall, these studies confirm that MLR is a simple, transparent, and effective method for estimating ET_0 across a wide range of climatic conditions. The inclusion of additional predictors such as solar radiation, humidity, and wind speed typically enhances model accuracy, yet reduced-variable models still provide useful predictions in regions with limited climate data.

This study focuses on the development and evaluation of MLR models for estimating ET_0 on an annual scale in the Darreh Dozdan River Basin of Iran. To date, no research has specifically applied regression-based approaches to estimate ET_0 in this basin, despite the importance of reliable evapotranspiration estimates for irrigation planning and water resource management. The study is motivated by the challenge of limited meteorological data, which often restricts the application of data-intensive methods such as the FAO-56 Penman–Monteith equation. The main objective is therefore to examine how effectively MLR can estimate ET_0 using reduced climate inputs and to identify the most influential variables contributing to model performance. By addressing this gap, the findings are expected to provide practical guidance for water managers and agricultural practitioners in data-scarce environments.

2. Materials and Methods

2.1. Study area

The Darreh Dozdan River (DDR) is one of the rivers of Lorestan Province, Iran. The DDR is one of the rivers that flows in the second-level watershed called Karkheh (Research Office of the Iran Water Resources Management Company, 2012). In the DDR basin, there are two stations: the Kuhdasht synoptic station and the Tange Siab hydrometric station. Both stations are located in Kuhdasht County. The Tange Siab station lacks the data required to estimate ET_0 in the DDR basin. In this study, the ET_0 in the DDR basin was estimated using meteorological data from the Kuhdasht synoptic Station during the period 1998–2022 (25 years). The Kuhdasht synoptic station has been established and operated by the Iran Meteorological Organization since 1997. It is situated at longitude 47°38′52″E, latitude 33°31′27″N, and an elevation of 1197 meters above sea level. Figure 1 shows the geographic location of the study area within Lorestan Province and Iran.

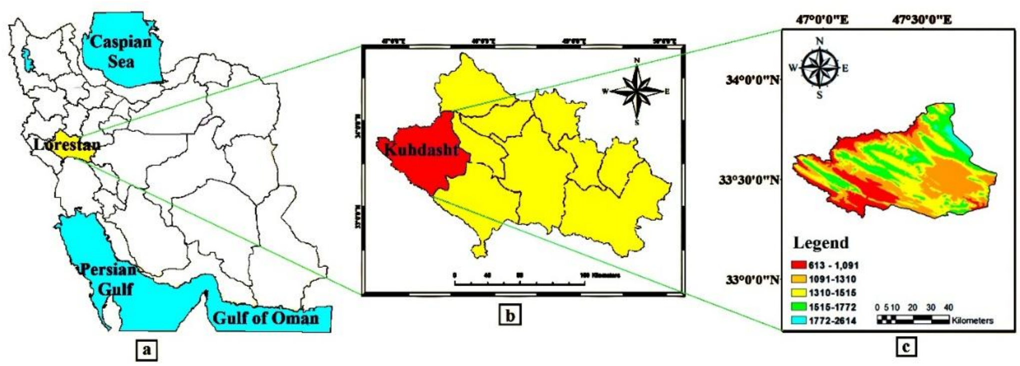

Fig. 1. a) Map of Iran, b) Map of Lorestan Province, c) Digital elevation model map of the study area.

2.2. FAO-56 Penman-Monteith method

In this study, ET_0 was calculated using the FAO-56 Penman-Monteith method through the CROPWAT 8.0 software. Reference crop evapotranspiration is defined as the evapotranspiration from an unlimited area of grass with a uniform height of 8-15 cm, with active growth without any water scarcity, nutrients, air, pests, and diseases that cover the entire ground surface with its shade (Ahmadpari and Khaustov, 2025b). The Food and Agriculture Organization of the United Nations (FAO) has recommended the FAO-56 Penman-Monteith method as the standard method for determining ET_0 (Ahmadpari et al., 2017). The reference crop is referred to by FAO as having 12 cm height and surface resistance of 70 s/m with Albedo coefficient-light reflection capability (Ahmadpari and Khaustov, 2025b). The value of the albedo coefficient is 0.23 (Ahmadpari et al., 2019). This method is expressed as Eq. 1 (Alazba et al., 2025).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\left[\frac{900}{(T+273)}\right]U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

where, ET_0 is the reference evapotranspiration (mm/day), R_n is the net radiation at the plant surface (MJ/m²/day), G is the soil heat flux (MJ/m²/day), T is the air temperature at a height of 2m (°C), U_2 is the wind speed at a height of 2m above ground (m/s), $(e_s - e_a)$ is the air saturation vapor pressure deficiency (kPa), Δ is the slope of the vapor pressure curve (kPa/°C), and γ is the hygrometric constant (kPa/°C). CROPWAT is a simulation model for climate, effective rainfall, crop, soil, water, and irrigation requirements, and irrigation scheduling (Binesh et al., 2020).

2.3. Multiple linear regression models

A multiple linear regression model was employed using the Regression option from the Analysis ToolPak of Microsoft Excel 2019 to quantify the relationship between ET_0 and the climatic variables. The regression model is expressed in Eq. 2 (Dimitriadou and Nikolakopoulos, 2022).

$$ET_0 = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 \quad (2)$$

where b_0 is the intercept, b_i are the regression coefficients, and x_1, x_2, x_3, x_4, x_5 are the independent variables representing minimum temperature, maximum temperature, relative humidity, wind speed, and sunshine hours, respectively. Considering the five independent variables, different regression models can be constructed by selecting subsets of these predictors. The total number of possible regression models that can be formed from k predictor variables is expressed in Equation 3 (Brooks and Ruengvirayudh, 2016).

$$Number\ of\ models = 2^k - 1 \quad (3)$$

The total number of regression models that can be formed from a set of k independent variables can be calculated using combinatorial methods. Specifically, the number of models that include exactly r predictor variables is expressed in Equation 4 (Stanley, 2011).

$$Number\ of\ models\ with\ r\ predictors = \binom{k}{r} = \frac{k!}{r!\,(k-r)!} \quad (4)$$

Where k is the total number of independent variables, r is the number of predictors included in the model, and ! denotes the factorial operation. In total, 31 possible regression models exist. Among these, there are five simple linear regression models, each including only one predictor variable. 10 regression models include two predictor variables, ten models that include three predictor variables, five models that include four predictor variables, and one model that includes all five predictor variables, which corresponds to the full model shown in Eq. 2.

2.4. Evaluation indices of models

2.4.1. Pearson correlation coefficient

The Pearson correlation coefficient (r), Eq. 5, measures the strength and direction of a linear relationship between two continuous variables, ranging from -1 (perfect negative) to 1 (perfect positive), with 0 indicating no correlation (Jiang and Sun, 2025).

$$Corr(X, Y) = \frac{Cov(X, Y)}{6_X 6_Y} \quad (5)$$

Where X and Y are the values of two variables, σ_X is the standard deviation of variable X, σ_Y is standard deviation of variable Y, Cov (X, Y) is the covariance between X and Y. Pearson's correlation coefficient is represented by r. The comparison of the Pearson correlation coefficient and correlation strength can be found in Table 1 (Jiang and Sun, 2025).

Table 1. Pearson correlation strength table (Jiang and Sun, 2025)

Correlation Coefficient (r)	Correlation Strength
0.0–0.2	Extremely weak correlated
0.2–0.4	Weak correlated
0.4–0.6	Medium correlated
0.6–0.8	Strong correlated
0.8–1.0	Highly correlated

2.4.2. Coefficient of determination

The coefficient of determination (R^2) measures how well independent variables explain the variance of the dependent variable, ranging from 0 to 1. Values above 0.6 indicate a strong explanatory power (Amini et al., 2019). It is calculated using Eq. 6 (Ahmadpari and Khaustov, 2025a).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(o_i - \bar{o})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2 \times \sum_{i=1}^{n}(s_i - \bar{s})^2}}\right)^2 \quad (6)$$

Where, s_i is the predicted value, o_i the observed value, $\bar{s}$ and $\bar{o}$ are predicted and observed average values, and n is the number of data points.

2.4.3. Adjusted R-Square

The Adjusted R^2 is a statistical measure used in regression analysis. It’s a modified version of the R-Square (R^2) that adjusts for the number of independent variables in the model. The problem with plain R^2 is that it always increases (or stays the same) when more variables are added, even if those variables don't actually improve the model. To fix this, Adjusted R^2 penalizes unnecessary predictors, ensuring only useful variables increase the score. The formula for calculating the Adjusted R^2 is Equation 7 (Karch, 2020).

$$\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1} \quad (7)$$

where n is the number of observations, p is the number of predictors (independent variables), and R^2 is the coefficient of determination.

2.4.4. Mean Absolute Error

Mean Absolute Error (MAE) is a metric used to evaluate the accuracy of a predictive model. It measures the average absolute difference between the predicted values and the actual observed values, without considering the direction (i.e., positive or negative). The formula for calculating the Mean Absolute Error is Eq. 8 (Amini et al., 2017; Ababakr et al., 2023).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |P_i - O_i| \quad (8)$$

Where n is the number of observations, P_i is the predicted values, and O_i is the actual (observed) values.

2.4.5. Root Mean Square Error

Root Mean Square Error (RMSE) is one of the most commonly used metrics to evaluate the accuracy of regression and forecasting models. It measures the square root of the average squared differences between the predicted values and the actual observed values. The formula for calculating the Root Mean Square Error is Eq. 9 (Amini 2020).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (9)$$

Where n is the number of observations, P_i is the predicted values, and O_i is the actual (observed) values.

2.4.6. Normalized Root Mean Square Error

The Normalized Root Mean Square Error (NRMSE) is a scaled version of the Root Mean Square Error (RMSE). This normalization helps interpret the error as a relative measure (percentage of the scale of the data). The NRMSE normalized by the mean is a metric used to assess the accuracy of predictive models. The formula for calculating the NRMSE is Eq. 10 (Dehghani et al., 2024).

$$\text{NRMSE} = \frac{\text{RMSE}}{\overline{\text{O}}} \quad (10)$$

Where RMSE is Root Mean Square Error, and $\overline{\text{O}}$ is the mean of the observed values.

2.4.7. Regression Standard Error

In multiple linear regression, the regression standard error (SE) has a specific meaning related to how well the regression model fits the observed data. Standard Error is the standard deviation of the residuals (errors), where residuals are the differences between observed values of the dependent variable and the values predicted by the regression model. It measures, on average, how far the observed data points fall from the regression hyperplane created by the multiple predictors. The formula for calculating the Regression Standard Error is Eq. 11 (Cochran, 1934).

$$SE = \sqrt{\frac{SSE}{n-k-1}} \quad (11)$$

Where SSE is the sum of squared residuals, n is the number of observations, and k is the number of predictors. A smaller standard error of the estimate indicates a better fit since observed values are closer to predicted values.

2.5. Analysis of Variance

In order to evaluate the overall significance of the MLR model, an analysis of variance table was employed. The analysis of variance table partitions the total variation in the dependent variable into two components: the variation explained by the regression model and the unexplained (residual) variation (Kim, 2014). The regression sum of squares represents the portion of variance accounted for by the independent variables, whereas the residual sum of squares corresponds to the variance that remains unexplained. Mean squares are obtained by dividing each sum of squares by its respective degrees of freedom (Miller et al., 2002). The ratio of the mean square for regression to the mean square for error yields the F-statistic, which tests the null hypothesis that all regression coefficients are equal to zero (i.e., the model has no explanatory power) (Das et al., 2022). A statistically significant F-test (< 0.05) indicates that the regression model explains a significant proportion of the variance in the dependent variable, thereby justifying further interpretation of the estimated coefficients (Sureiman and Mangera, 2020).

2.6. Estimation and Significance Testing of Regression Coefficients

The regression coefficients were estimated by the ordinary least squares method, which minimizes the sum of squared residuals to produce unbiased and efficient estimates under the classical assumptions of linear regression (Draper and Smith, 1998). This method is widely used in environmental and hydrological studies due to its simplicity and interpretability (Montgomery et al., 2021). To assess the reliability and significance of regression coefficient estimates, the standard error, t statistic, and p-value statistical parameters were calculated.

2.6.1. Standard Error

Standard Error (SE) represents the estimated standard deviation of the sampling distribution of a regression coefficient. It measures the precision of the coefficient estimate; a smaller Standard Error indicates higher precision. The Standard Error is calculated as the square root of the variance of the coefficient estimate, which depends on the residual variance and the independent variables matrix. The standard error of the estimated regression coefficient $\hat{\beta}_i$, denoted as $SE(\hat{\beta}_i)$, measures the precision of the coefficient estimate and reflects the variability of the estimator across different samples. The formula for calculating the regression coefficient Standard Error is Equation 12 (Draper and Smith, 1998).

$$SE(\hat{\beta}_i) = \sqrt{\hat{\sigma}^2 \times (X^T X)^{-1}_{ii}} \quad (12)$$

Where, $\hat{\sigma}^2$ is the estimated variance of the residuals, and $(X^T X)^{-1}_{ii}$ is the i-th diagonal element of the inverse of the matrix of predictors.

2.6.2. t Statistic

t Statistic (t-stat) is calculated as the ratio of the estimated coefficient $\hat{\beta}_i$, to its standard error. The formula for calculating the t-statistic is Equation 13 (Walpole et al., 2022).

$$t_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (12)$$

This statistic tests the null hypothesis that the true coefficient is zero (no effect). A larger absolute value of t_i implies stronger evidence against the null hypothesis (Walpole et al., 2022).

2.6.3. P-value

P-value is derived from the t-distribution with degrees of freedom equal to $n - p - 1$, where n is the sample size, and p is the number of predictors. It represents the probability of obtaining a t_i as extreme as observed, assuming the null hypothesis is true. A commonly used threshold for significance is $p < 0.05$. Lower P-values indicate statistically significant predictors (Montgomery et al., 2021).

3. Results and Discussion

3.1. Simple linear regression models

Table 2 presents the regression coefficients, standard errors, t-statistics, and p-values for five simple linear models linking each meteorological variable to reference evapotranspiration.

Table 2. Summary of regression coefficients and statistical significance for the simple linear regression models

Models	Variable	Coefficients	Standard Error	t Stat	P-value
MLR1	Intercept	0.98	0.78	1.26	0.22
	Minimum temperature	0.40	0.11	3.83	0.00
MLR2	Intercept	-7.34	2.92	-2.51	0.02
	Maximum temperature	0.45	0.12	3.86	0.00
MLR3	Intercept	6.19	1.41	4.39	0.00
	Relative humidity	-0.05	0.03	-1.59	0.12
MLR4	Intercept	2.33	0.10	22.61	0.00
	Wind speed	0.01	0.00	16.33	0.00
MLR5	Intercept	3.21	2.68	1.20	0.24
	Sunshine	0.08	0.30	0.27	0.79

The results indicate that minimum temperature, maximum temperature, and wind speed have significant positive effects on ET_0 (all with $p < 0.01$), whereas relative humidity has a negative but non-significant coefficient ($p = 0.12$), and sunshine hours have no statistically significant effect ($p = 0.79$). In practical terms, rising minimum or maximum temperature leads to higher ET_0, and even modest increases in wind speed are associated with increased evaporative demand, while humidity and sunshine alone do not explain much of the variability in ET_0 in this dataset. These findings align with many recent studies. Taheri et al. (2025) demonstrated that, among the input variables in deep learning models, radiation had the largest influence on evapotranspiration at approximately 42%, followed by maximum temperature at around 32%, while relative humidity had the smallest contribution. Table 3 presents the performance of the simple linear regression models for estimating ET_0 using individual meteorological variables.

Table 3. Performance of simple linear regression models

Models	r	R^2	Adjusted R^2	SE	MAE	RMSE	NRMSE
MLR1	0.62	0.39	0.36	0.39	0.32	0.37	0.09
MLR2	0.63	0.39	0.37	0.39	0.32	0.37	0.09
MLR3	0.32	0.10	0.06	0.47	0.40	0.45	0.11
MLR4	0.96	0.92	0.92	0.14	0.11	0.13	0.03
MLR5	0.06	0.00	-0.04	0.50	0.40	0.48	0.12

In this analysis, r indicates the strength of the linear association, while the R^2 shows the proportion of ET_0 variability explained by each predictor. According to the correlation strength classification (Table 1), wind speed (MLR4) exhibited a high correlation with ET_0 ($r = 0.96$), accounting for 92% of the variance ($R^2 = 0.92$) and achieving the lowest error indices (RMSE = 0.13, NRMSE = 0.03). Minimum and maximum temperatures (MLR1 and MLR2) demonstrated strong correlations ($r \approx 0.62–0.63$) and explained about 39% of the variance in ET_0 ($R^2 = 0.39$), with moderate levels of prediction error (RMSE = 0.37, NRMSE = 0.09). In contrast, relative humidity (MLR3) showed only a weak correlation ($r = 0.32$), with a limited explanatory power ($R^2 = 0.10$). Sunshine duration (MLR5) exhibited an extremely weak correlation ($r = 0.06$) and virtually no explanatory capability ($R^2 \approx 0$), coupled with the highest error values. These findings demonstrate that wind speed is the dominant climatic factor regulating ET_0 in the study area, followed by air temperature, whereas relative humidity and sunshine hours provide minimal contribution when considered individually.

Figure 2 illustrates the comparison between ET_0 values estimated by the FAO-56 Penman–Monteith method, considered as the standard approach, and those predicted by the simple linear regression models.

The visual patterns confirm the statistical findings reported in Table 3: wind speed (MLR4) shows the closest agreement with the standard method, reflecting its high correlation ($r = 0.96$) and strong explanatory power ($R^2 = 0.92$). Minimum and maximum temperatures (MLR1 and MLR2) also provide reasonable estimates with moderate agreement, consistent with their strong but lower correlations. By contrast, the models based on relative humidity (MLR3) and sunshine duration (MLR5) deviate considerably from the FAO-56 Penman–Monteith estimates, which supports their weak or negligible predictive ability. Overall, Fig. 2 visually reinforces the conclusion that wind speed is the dominant climatic driver of ET_0 in the study area, followed by air temperature, while relative humidity and sunshine contribute little when applied individually.

Table 4 presents the results of the analysis of variance (ANOVA) conducted for the simple linear regression models developed to estimate reference evapotranspiration

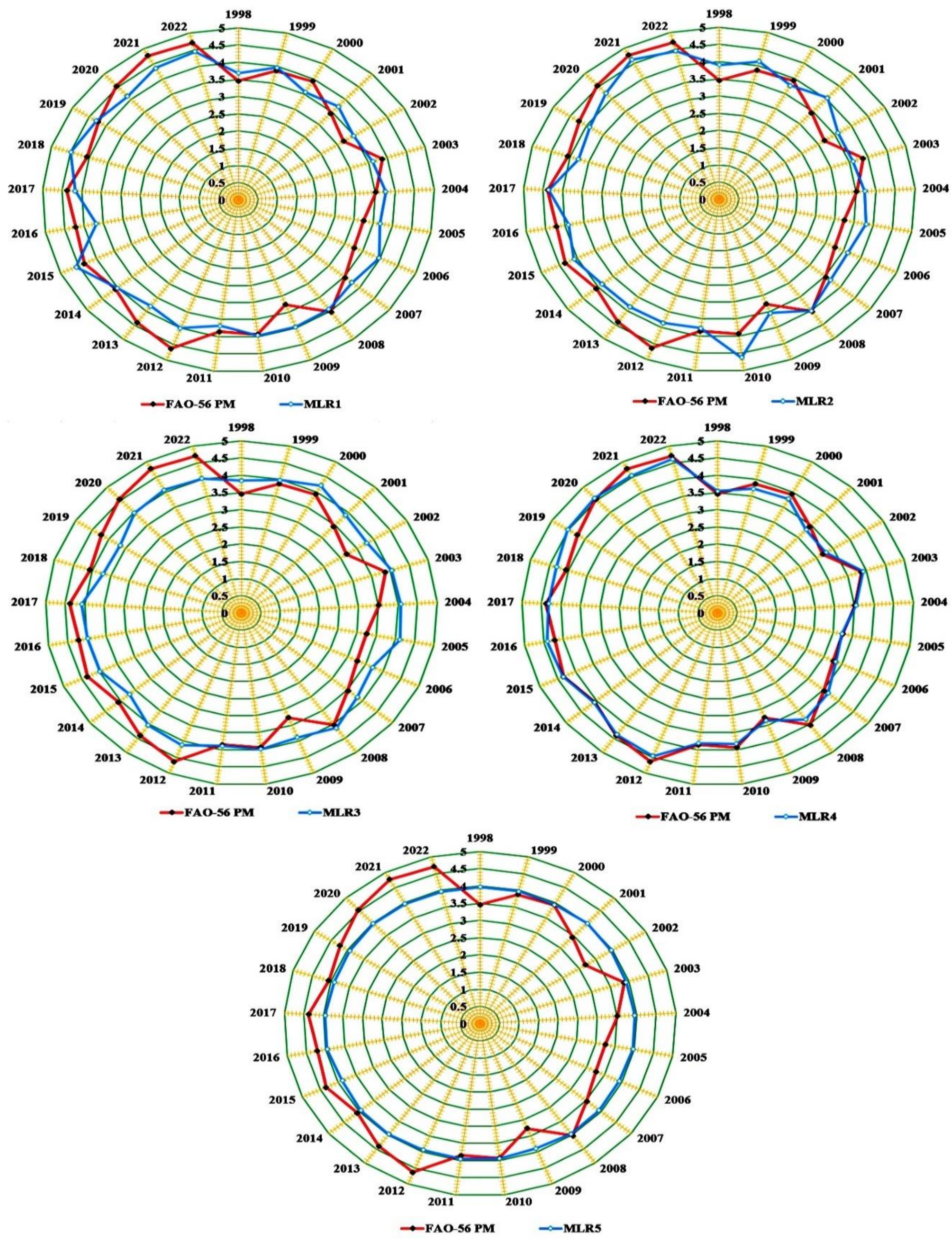

Fig. 2. Comparison of ET_0 Estimated by the FAO-56 Penman-Monteith Method and Simple Linear Regression Models

.

Table 4. ANOVA for simple linear regression models

Models	Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F Value	Significance F
MLR1	Regression	1	2.22	2.22	14.68	0.00
	Residual	23	3.48	0.15		
	Total	24	5.70			
MLR2	Regression	1	2.24	2.24	14.93	0.00
	Residual	23	3.46	0.15		
	Total	24	5.70			
MLR3	Regression	1	0.57	0.57	2.54	0.12
	Residual	23	5.13	0.22		
	Total	24	5.70			
MLR4	Regression	1	5.25	5.25	266.65	0.00
	Residual	23	0.45	0.02		
	Total	24	5.70			
MLR5	Regression	1	0.02	0.02	0.08	0.79
	Residual	23	5.68	0.25		
	Total	24	5.70			

The ANOVA test evaluates whether the regression model as a whole significantly explains the variability in ET_0. The results indicate that models based on minimum temperature (MLR1), maximum temperature (MLR2), and wind speed (MLR4) are statistically significant (Significance $F < 0.01$), confirming that these predictors provide meaningful contributions to the estimation of ET_0. Among them, wind speed (MLR4) shows the strongest performance with an extremely high F value (266.65), highlighting its dominant role. By contrast, relative humidity (MLR3) and sunshine duration (MLR5) are not statistically significant (Significance $F > 0.05$), indicating that they do not offer reliable explanatory power for ET_0 when considered individually. These findings are fully consistent with the correlation and model performance results presented in Tables 2 and 3, reinforcing the conclusion that wind speed and air temperature are the most influential climatic variables governing ET_0 in the study area.

3.2. Two predictor variables

Table 5 summarizes the regression coefficients, standard errors, t-statistics, and significance levels for ten MLR models that incorporate two meteorological variables as predictors of reference evapotranspiration. The results demonstrate that models combining wind speed with either maximum temperature (MLR7), relative humidity (MLR13), or minimum temperature (MLR15) perform particularly well, as both wind speed and the accompanying variable are statistically significant ($p < 0.05$), with especially strong effects observed for wind speed ($t > 12$ across models). For example, in MLR7, both maximum temperature ($p < 0.01$) and wind speed ($p < 0.01$) contribute positively to ET_0, while in MLR13, wind speed ($p < 0.01$) and relative humidity ($p < 0.01$) jointly explain variations, with humidity showing a negative influence. This study demonstrated that multivariable regression models, particularly those including wind speed, achieved high accuracy in predicting reference evapotranspiration. This finding is consistent with recent research using Long Short-Term Memory models, where combining temperature and wind speed also provided reliable ET_0 predictions with R^2 above 0.75 (Karuppanan et al., 2025). Models combining temperature with sunshine, such as MLR6 and MLR10, highlight the importance of temperature, whereas sunshine shows no significant or only marginal effects ($p > 0.05$). Similarly, MLR11 indicates that neither relative humidity nor sunshine alone provides a reliable contribution to ET_0 estimation. Overall, these results suggest that while temperature variables consistently show significant positive effects, the inclusion of wind speed markedly strengthens model performance, confirming its dominant role in controlling ET_0, whereas sunshine hours have little explanatory power even in multivariate combinations

Table 5. Summary of regression coefficients and statistical significance for ten MLR models with two independent variables

Models	Variable	Coefficients	Standard Error	t Stat	P-value
MLR6	Intercept	-6.08	3.09	-1.97	0.06
	Maximum temperature	0.50	0.12	4.05	0.00
	Sunshine	-0.29	0.25	-1.17	0.26
MLR7	Intercept	-0.94	0.97	-0.97	0.34
	Maximum temperature	0.14	0.04	3.38	0.00
	Wind speed	0.01	0.00	15.27	0.00
MLR8	Intercept	-5.66	3.67	-1.54	0.14
	Maximum temperature	0.42	0.12	3.40	0.00
	Relative humidity	-0.02	0.03	-0.77	0.45
MLR9	Intercept	3.66	1.13	3.23	0.00
	Minimum temperature	0.44	0.09	4.82	0.00
	Relative humidity	-0.06	0.02	-2.94	0.01
MLR10	Intercept	-3.76	2.51	-1.50	0.15
	Minimum temperature	0.48	0.11	4.50	0.00
	Sunshine	0.47	0.24	1.97	0.06
MLR11	Intercept	6.43	3.35	1.92	0.07
	Relative humidity	-0.05	0.03	-1.53	0.14
	Sunshine	-0.02	0.30	-0.08	0.94
MLR12	Intercept	1.05	0.74	1.42	0.17
	Wind speed	0.01	0.00	17.11	0.00
	Sunshine	0.14	0.08	1.76	0.09
MLR13	Intercept	3.78	0.33	11.64	0.00
	Relative humidity	-0.03	0.01	-4.60	0.00
	Wind speed	0.01	0.00	21.63	0.00
MLR14	Intercept	-5.40	2.81	-1.92	0.07
	Minimum temperature	0.26	0.11	2.31	0.03
	Maximum temperature	0.30	0.13	2.35	0.03
MLR15	Intercept	2.37	0.31	7.66	0.00
	Minimum temperature	-0.01	0.05	-0.13	0.90
	Wind speed	0.01	0.00	12.13	0.00

Table 6. Performance of regression models with two independent variables

Models	r	R^2	Adjusted R^2	SE	MAE	RMSE	NRMSE
MLR6	0.66	0.43	0.38	0.38	0.30	0.36	0.09
MLR7	0.97	0.95	0.94	0.12	0.09	0.11	0.03
MLR8	0.64	0.41	0.36	0.39	0.31	0.37	0.09
MLR9	0.75	0.56	0.52	0.34	0.26	0.32	0.08
MLR10	0.69	0.48	0.43	0.37	0.29	0.34	0.09
MLR11	0.32	0.10	0.02	0.48	0.40	0.45	0.11
MLR12	0.96	0.93	0.92	0.13	0.10	0.13	0.03
MLR13	0.98	0.96	0.96	0.10	0.08	0.10	0.03
MLR14	0.72	0.51	0.47	0.36	0.28	0.33	0.08
MLR15	0.96	0.92	0.91	0.14	0.11	0.13	0.03

The results clearly demonstrate that models incorporating wind speed consistently outperform others. In particular, MLR13 (relative humidity + wind speed) achieved the best performance, with a very high correlation ($r = 0.98$), the highest explanatory power ($R^2 = 0.96$), and the lowest error values (RMSE = 0.10, NRMSE = 0.03). Similarly, MLR7 (maximum temperature + wind speed), MLR12 (wind speed + sunshine), and MLR15 (minimum temperature + wind speed) also exhibited excellent predictive ability (R^2 = 0.92–0.95, RMSE = 0.11–0.13), confirming the dominant role of wind speed in ET_0 estimation when combined with other variables. By contrast, models without wind speed, such as MLR6 (maximum temperature + sunshine), MLR8 (maximum temperature + relative humidity), MLR9 (minimum temperature

+ relative humidity), MLR10 (minimum temperature + sunshine), and especially MLR11 (relative humidity + sunshine), performed markedly worse, with much lower correlations (r = 0.32–0.75) and higher errors. Notably, MLR11 explained only about 10% of the variance (R^2 = 0.10) and had the poorest accuracy (RMSE = 0.45, NRMSE = 0.11). Overall, the results highlight that the integration of wind speed with another climatic factor, particularly relative humidity or temperature, provides the most robust and accurate models for estimating ET_0, while models based on sunshine and humidity alone are inadequate. Figure 3 illustrates the comparison between ET_0 values estimated by the FAO-56 Penman–Monteith method and those predicted by MLR models incorporating two independent variables.

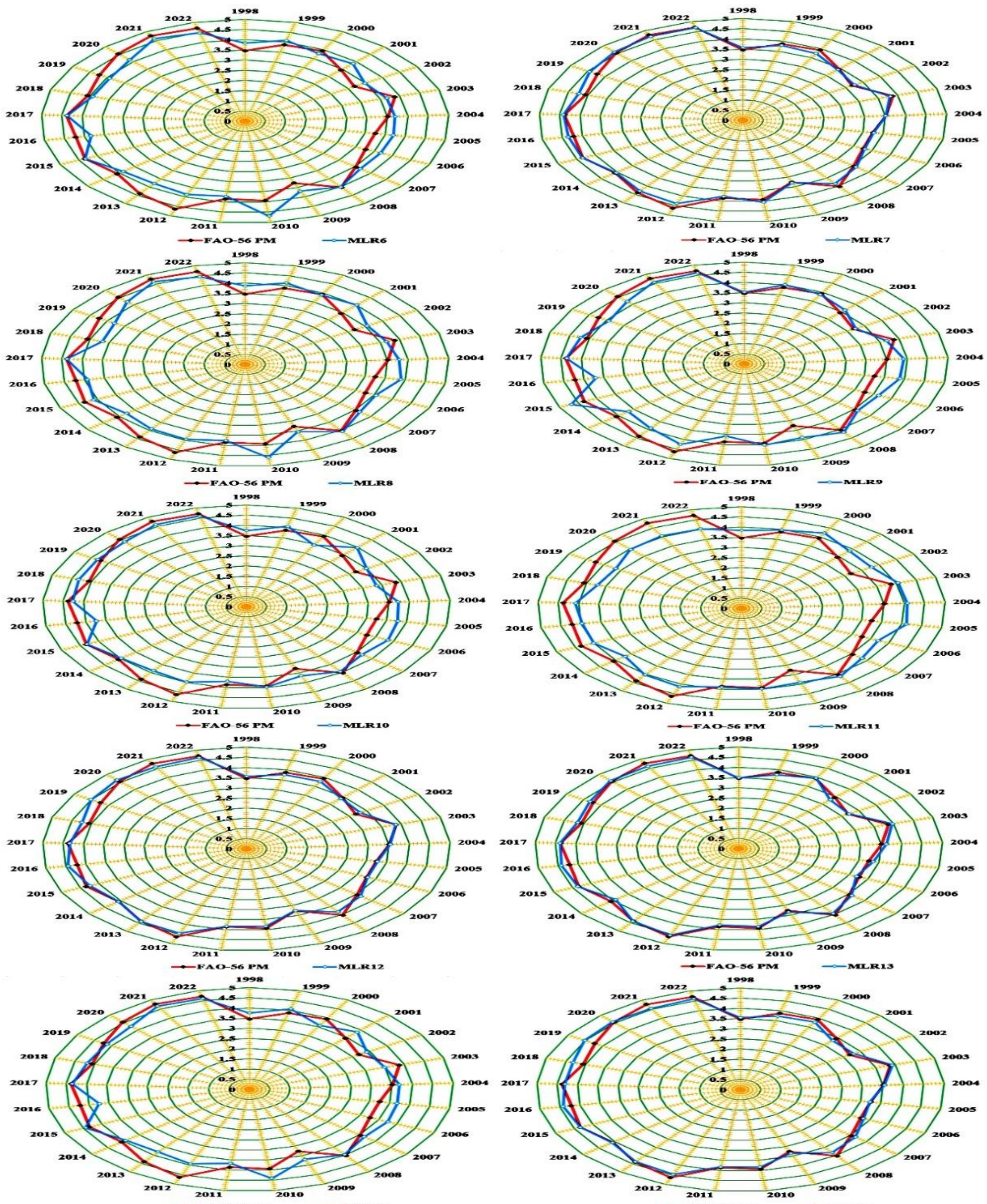

Fig. 3. Comparison of ET_0 Estimated by the FAO-56 Penman-Monteith Method and Regression Models with Two Predictor Variables

Figure 3 provides a visual assessment of model accuracy, highlighting the degree of agreement between observed ET_0 and estimates from different model combinations. Consistent with Table 6, models that include wind speed, particularly MLR7, MLR12, MLR13, and MLR15, show the closest alignment with the standard FAO-56 estimates, while models without wind speed exhibit larger deviations. This visual comparison reinforces the dominant influence of wind speed on ET_0 estimation and confirms the superior performance of multivariate models that combine wind speed with either temperature or relative humidity. Table 7 presents the results of the ANOVA for MLR models incorporating two independent variables to predict reference evapotranspiration.

Table 7. ANOVA for regression models with two independent variables

Models	Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F Value	Significance F
MLR6	Regression	2	2.45	1.22	8.27	0.00
	Residual	22	3.25	0.15		
	Total	24	5.70			
MLR7	Regression	2	5.40	2.70	199.34	0.00
	Residual	22	0.30	0.01		
	Total	24	5.70			
MLR8	Regression	2	2.33	1.17	7.63	0.00
	Residual	22	3.37	0.15		
	Total	24	5.70			
MLR9	Regression	2	3.20	1.60	14.09	0.00
	Residual	22	2.50	0.11		
	Total	24	5.70			
MLR10	Regression	2	2.74	1.37	10.21	0.00
	Residual	22	2.96	0.13		
	Total	24	5.70			
MLR11	Regression	2	0.57	0.28	1.22	0.32
	Residual	22	5.13	0.23		
	Total	24	5.70			
MLR12	Regression	2	5.30	2.65	146.98	0.00
	Residual	22	0.40	0.02		
	Total	24	5.70			
MLR13	Regression	2	5.47	2.74	260.83	0.00
	Residual	22	0.23	0.01		
	Total	24	5.70			
MLR14	Regression	2	2.92	1.46	11.53	0.00
	Residual	22	2.78	0.13		
	Total	24	5.70			
MLR15	Regression	2	5.25	2.62	127.64	0.00
	Residual	22	0.45	0.02		
	Total	24	5.70			

The results indicate that models including wind speed (MLR7, MLR12, MLR13, and MLR15) are highly significant (Significance $F < 0.01$) and exhibit very large F values (ranging from 127.64 to 260.83), confirming their strong explanatory power. Models combining temperature with other variables, such as MLR6, MLR8, MLR9, MLR10, and MLR14 are also significant (Significance $F < 0.01$), though with lower F values, suggesting moderate explanatory ability. By contrast, MLR11, which includes only relative humidity and sunshine, is not statistically significant (Significance $F = 0.32$), indicating that these variables together do not reliably account for ET_0 variability. Overall, the ANOVA results reinforce the conclusions from Tables 5 and 6: wind speed is the dominant driver of ET_0, and models that include it alongside temperature or humidity provide the most accurate and reliable predictions.

3.3. Three predictor variables

Table 8 presents the regression coefficients, standard errors, t-statistics, and significance levels for ten MLR models incorporating three independent meteorological variables to predict reference evapotranspiration.

Table 8. Summary of regression coefficients and statistical significance for ten MLR models with three independent variables

Models	Variable	Coefficients	Standard Error	t Stat	P-value
MLR16	Intercept	0.58	1.00	0.58	0.57
	Minimum temperature	0.04	0.05	0.70	0.49
	Wind speed	0.01	0.00	11.79	0.00
	Sunshine	0.17	0.09	1.87	0.08
MLR17	Intercept	1.20	0.88	1.36	0.19
	Maximum temperature	0.10	0.03	3.08	0.01
	Relative humidity	-0.02	0.01	-4.29	0.00
	Wind speed	0.01	0.00	20.59	0.00
MLR18	Intercept	-3.89	3.88	-1.00	0.33
	Maximum temperature	0.47	0.13	3.68	0.00
	Relative humidity	-0.02	0.03	-0.94	0.36
	Sunshine	-0.32	0.25	-1.28	0.21
MLR19	Intercept	-1.06	1.01	-1.05	0.31
	Maximum temperature	0.13	0.05	2.71	0.01
	Wind speed	0.01	0.00	14.54	0.00
	Sunshine	0.04	0.08	0.54	0.59
MLR20	Intercept	-0.26	2.58	-0.10	0.92
	Minimum temperature	0.50	0.09	5.29	0.00
	Relative humidity	-0.06	0.02	-2.68	0.01
	Sunshine	0.36	0.21	1.68	0.11
MLR21	Intercept	3.63	0.34	10.73	0.00
	Minimum temperature	0.05	0.04	1.34	0.19
	Relative humidity	-0.03	0.01	-4.87	0.00
	Wind speed	0.01	0.00	15.03	0.00
MLR22	Intercept	-5.70	2.92	-1.95	0.06
	Minimum temperature	0.32	0.16	1.97	0.06
	Maximum temperature	0.23	0.18	1.26	0.22
	Sunshine	0.17	0.33	0.51	0.61
MLR23	Intercept	-1.02	0.96	-1.06	0.30
	Minimum temperature	-0.05	0.04	-1.27	0.22
	Maximum temperature	0.15	0.04	3.65	0.00
	Wind speed	0.01	0.00	13.79	0.00
MLR24	Intercept	-0.57	3.53	-0.16	0.87
	Minimum temperature	0.35	0.12	3.07	0.01
	Maximum temperature	0.17	0.13	1.26	0.22
	Relative humidity	-0.05	0.02	-2.04	0.05
MLR25	Intercept	2.98	0.72	4.15	0.00
	Relative humidity	-0.03	0.01	-4.23	0.00
	Wind speed	0.01	0.00	21.94	0.00
	Sunshine	0.08	0.06	1.26	0.22

Table 8 shows that wind speed consistently exhibits highly significant positive effects ($p < 0.01$) across most models, confirming its dominant influence on ET_0. Temperature variables (minimum or maximum) also generally have significant positive coefficients, although their significance varies depending on the combination of other predictors. Relative humidity frequently has a negative effect, with significance in several models (e.g., MLR17, MLR20, MLR21, MLR24, MLR25), indicating its dampening influence on ET_0. Sunshine hours appear to have a limited or marginal impact, with p-values mostly above 0.05, suggesting that this variable contributes little to ET_0 prediction when combined with other climatic factors. Overall, these findings reinforce the conclusions drawn from simpler models: while temperature and relative humidity can influence ET_0, wind speed remains the most robust predictor, and including it alongside other variables consistently improves model performance. The results of this study are inconsistent with those of Koç and Can (2023). Koç et al. (2023) reported that MLR models incorporating solar radiation and sunshine hours data estimated daily ET_0 more accurately than those using other variables. They found that sunshine hours had a substantial effect on ET_0 estimation, with RMSE values ranging between 0.457 and 0.750 mm/day. Table 9 summarizes the performance metrics of MLR models incorporating three independent variables to estimate reference evapotranspiration.

Table 9. Performance of regression models with three independent variables

Models	r	R^2	Adjusted R^2	SE	MAE	RMSE	NRMSE
MLR16	0.97	0.93	0.92	0.14	0.10	0.12	0.03
MLR17	0.99	0.97	0.97	0.09	0.07	0.08	0.02
MLR18	0.67	0.45	0.37	0.39	0.29	0.35	0.09
MLR19	0.97	0.95	0.94	0.12	0.08	0.11	0.03
MLR20	0.78	0.61	0.56	0.32	0.24	0.30	0.08
MLR21	0.98	0.96	0.96	0.10	0.08	0.09	0.02
MLR22	0.72	0.52	0.45	0.36	0.27	0.33	0.08
MLR23	0.98	0.95	0.94	0.11	0.08	0.11	0.03
MLR24	0.77	0.59	0.53	0.33	0.26	0.30	0.08
MLR25	0.98	0.96	0.96	0.10	0.08	0.09	0.02

The results indicate that models including wind speed consistently achieve the highest predictive accuracy. In particular, MLR17 (maximum temperature + relative humidity + wind speed), MLR21 (minimum temperature + relative humidity + wind speed), MLR23 (minimum temperature + maximum temperature + wind speed), and MLR25 (relative humidity + wind speed + sunshine) demonstrate excellent performance, with correlation coefficients (r) between 0.98–0.99, R^2 values of 0.95–0.97, and very low errors (RMSE = 0.08–0.12, NRMSE = 0.02–0.03). Models that do not include wind speed or have weaker variable combinations, such as MLR18 (maximum temperature + relative humidity + sunshine) and MLR22 (minimum temperature + maximum temperature + sunshine), show considerably lower performance (R^2 = 0.37–0.45, RMSE = 0.33–0.35). These results reinforce the critical role of wind speed in ET_0 estimation and suggest that incorporating it alongside temperature and relative humidity provides the most robust and accurate models, while combinations lacking wind speed are substantially less reliable.

Figure 4 presents a comparison between ET_0 estimated by the FAO-56 Penman–Monteith method and values predicted by MLR models incorporating three independent variables.

Figure 4 visually confirms the findings reported in Table 9: models that include wind speed in combination with temperature and/or relative humidity (e.g., MLR17, MLR21, MLR23, MLR25) show the closest agreement with the FAO-56 estimates, exhibiting minimal deviations. Models without wind speed, or with less optimal variable combinations, display greater discrepancies from the standard ET_0 values. Overall, Figure 4 reinforces that incorporating wind speed into multivariate regression models substantially improves ET_0 prediction accuracy, while the inclusion of sunshine alone contributes little when combined with other predictors. Table 10 presents the analysis of variance for MLR models incorporating three independent meteorological variables to predict reference evapotranspiration.

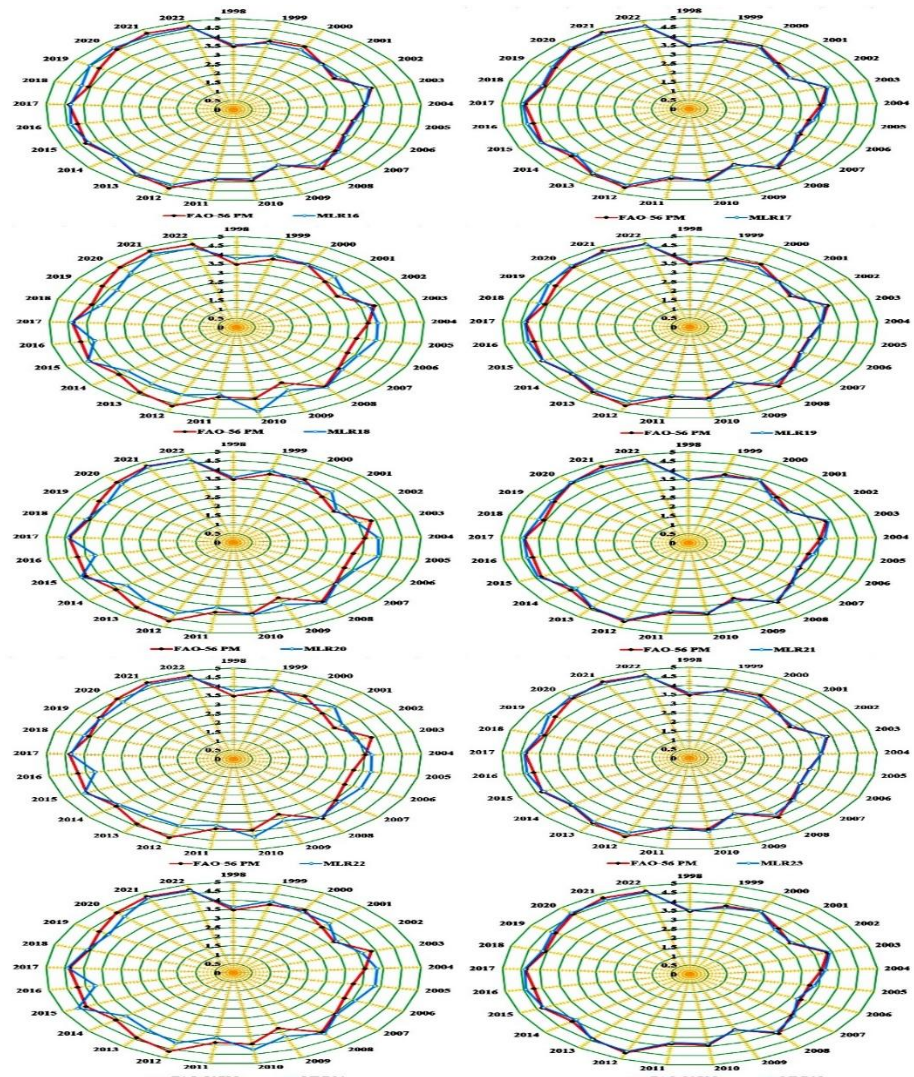

Fig. 4. Comparison of ET_0 Estimated by the FAO-56 Penman-Monteith Method and Regression Models with Three Predictor Variables

Table 10 shows that all ten models are statistically significant (Significance F = 0.00), with F values ranging from 5.78 to 244.09. Models including wind speed combined with temperature and/or relative humidity (e.g., MLR17, MLR21, MLR23, MLR25) exhibit the highest F values and lowest residual variance, confirming their superior explanatory power. Even models with less optimal combinations, such as MLR18 and MLR22, are statistically significant, although they have lower F values and higher residuals, indicating moderate predictive capability. Overall, the ANOVA results reinforce the findings from the regression coefficients and performance metrics: incorporating wind speed, especially alongside temperature and relative humidity, produces the most robust and accurate models for ET_0 estimation, while other variable combinations remain significant but less effective.

3.4. Four predictor variables

Table 11 presents the regression coefficients, standard errors, t-statistics, and significance levels for five MLR models developed using four independent meteorological variables to estimate reference evapotranspiration.

Table 10. ANOVA for regression models with three independent variables

Models	Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F Value	Significance F
MLR16	Regression	3	5.31	1.77	95.88	0.00
	Residual	21	0.39	0.02		
	Total	24	5.70			
MLR17	Regression	3	5.54	1.85	244.09	0.00
	Residual	21	0.16	0.01		
	Total	24	5.70			
MLR18	Regression	3	2.58	0.86	5.78	0.00
	Residual	21	3.12	0.15		
	Total	24	5.70			
MLR19	Regression	3	5.41	1.80	128.75	0.00
	Residual	21	0.29	0.01		
	Total	24	5.70			
MLR20	Regression	3	3.50	1.17	11.12	0.00
	Residual	21	2.20	0.10		
	Total	24	5.70			
MLR21	Regression	3	5.49	1.83	180.84	0.00
	Residual	21	0.21	0.01		
	Total	24	5.70			
MLR22	Regression	3	2.95	0.98	7.52	0.00
	Residual	21	2.75	0.13		
	Total	24	5.70			
MLR23	Regression	3	5.42	1.81	137.10	0.00
	Residual	21	0.28	0.01		
	Total	24	5.70			
MLR24	Regression	3	3.38	1.13	10.18	0.00
	Residual	21	2.32	0.11		
	Total	24	5.70			
MLR25	Regression	3	5.49	1.83	179.03	0.00
	Residual	21	0.21	0.01		
	Total	24	5.70			

The results highlight that wind speed remains the most consistent and highly significant predictor ($p < 0.01$) across almost all models, with strong positive effects on ET_0. The results of this study are inconsistent with those of Dimitriadou and Nikolakopoulos (2022), who reported that MLR models could not capture the complex relationship between wind speed and ET_0. Maximum temperature also shows significant positive effects in several models (MLR26, MLR28, MLR30), while minimum temperature is only significant in selected cases (MLR27 and MLR29). Relative humidity consistently exerts a negative and statistically significant effect in most models, confirming its dampening influence on ET_0. The results of this study are consistent with those of Purohit et al. (2016), who investigated the influence of various meteorological variables on evapotranspiration under the humid climate conditions of the Konkan region. Their results showed that relative humidity has a significant negative relationship with ET_0, where an increase in relative humidity causes a decrease in ET_0, while temperature and wind speed have positive effects. Sunshine appears to have a weaker and less consistent role: it is not significant in most models, except for a marginal effect in MLR29 ($p = 0.05$). Overall, these findings indicate that combining wind speed with temperature (minimum or maximum) and relative humidity provides the most reliable models for ET_0 prediction, while the addition of sunshine does not markedly improve predictive performance. Table 12 summarizes the performance indices of MLR models with four predictor variables for estimating reference evapotranspiration.

Table 11. Summary of regression coefficients and statistical significance for five MLR models with four independent variables

Models	Variable	Coefficients	Standard Error	t Stat	P-value
MLR26	Intercept	1.27	0.95	1.34	0.20
	Minimum temperature	0.01	0.04	0.23	0.82
	Maximum temperature	0.09	0.04	2.61	0.02
	Relative humidity	-0.03	0.01	-3.87	0.00
	Wind speed	0.01	0.00	16.52	0.00
MLR27	Intercept	-0.57	3.53	-0.16	0.87
	Minimum temperature	0.48	0.17	2.89	0.01
	Maximum temperature	0.03	0.19	0.13	0.90
	Relative humidity	-0.05	0.02	-2.23	0.04
	Sunshine	0.33	0.31	1.05	0.31
MLR28	Intercept	-0.92	1.01	-0.91	0.37
	Minimum temperature	-0.07	0.06	-1.19	0.25
	Maximum temperature	0.17	0.06	2.87	0.01
	Wind speed	0.01	0.00	13.42	0.00
	Sunshine	-0.05	0.11	-0.41	0.69
MLR29	Intercept	2.17	0.76	2.87	0.01
	Minimum temperature	0.08	0.04	2.17	0.04
	Relative humidity	-0.03	0.01	-4.97	0.00
	Wind speed	0.01	0.00	15.29	0.00
	Sunshine	0.13	0.06	2.12	0.05
MLR30	Intercept	1.16	0.93	1.24	0.23
	Maximum temperature	0.09	0.04	2.65	0.02
	Relative humidity	-0.02	0.01	-4.13	0.00
	Wind speed	0.01	0.00	19.33	0.00
	Sunshine	0.01	0.06	0.20	0.84

Table 12. Performance of regression models with four independent variables

Models	r	R^2	Adjusted R^2	SE	MAE	RMSE	NRMSE
MLR26	0.99	0.97	0.97	0.09	0.07	0.08	0.02
MLR27	0.78	0.61	0.54	0.33	0.24	0.30	0.08
MLR28	0.98	0.95	0.94	0.12	0.08	0.10	0.03
MLR29	0.98	0.97	0.96	0.09	0.07	0.08	0.02
MLR30	0.99	0.97	0.97	0.09	0.07	0.08	0.02

The results clearly demonstrate that most of the models achieve very high predictive accuracy, particularly MLR26, MLR29, and MLR30, which report correlation coefficients (r) of 0.98–0.99, R^2 values of 0.97, and very low error indices (RMSE = 0.08, NRMSE = 0.02). MLR28 also performs strongly, with r = 0.98 and R^2 = 0.95, though its accuracy is slightly lower compared to the top-performing models. In contrast, MLR27 shows considerably weaker performance, with r = 0.78, R^2 = 0.61, and relatively higher errors (RMSE = 0.30, NRMSE = 0.08), indicating that the specific variable combination in this model is less effective for ET_0 prediction. Overall, the results confirm that including wind speed, temperature, and relative humidity together yields highly reliable models, while the addition of sunshine improves prediction only marginally. Figure 5 illustrates the comparison between ET_0 estimated by the FAO-56 Penman–Monteith method and the predictions obtained from MLR models using four independent variables.

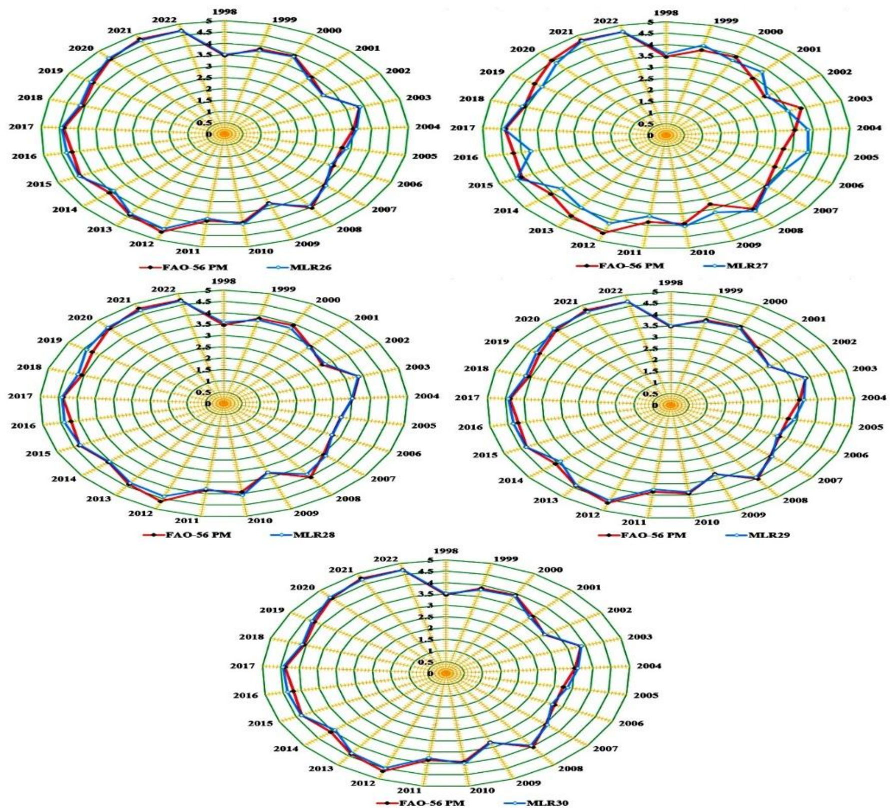

Fig. 5. Comparison of ET_0 Estimated by the FAO-56 Penman-Monteith Method and Regression Models with Four Predictor Variables

Figure 5 confirms the results presented in Table 12: models such as MLR26, MLR29, and MLR30 demonstrate very close alignment with the FAO-56 estimates, reflecting their high accuracy and low prediction errors. MLR28 also shows a strong agreement, although with slightly larger deviations. In contrast, MLR27 displays noticeable discrepancies from the standard values, consistent with its weaker statistical performance. Table 13 presents the ANOVA results for the MLR models with four independent variables.

Table 13. ANOVA for regression models with four independent variables

Models	Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F Value	Significance F
	Regression	4	5.54	1.39	174.83	0.00
MLR26	Residual	20	0.16	0.01		
	Total	24	5.70			
	Regression	4	3.50	0.87	7.95	0.00
MLR27	Residual	20	2.20	0.11		
	Total	24	5.70			
	Regression	4	5.43	1.36	98.79	0.00
MLR28	Residual	20	0.27	0.01		
	Total	24	5.70			
	Regression	4	5.53	1.38	159.24	0.00
MLR29	Residual	20	0.17	0.01		
	Total	24	5.70			
	Regression	4	5.54	1.39	174.71	0.00
MLR30	Residual	20	0.16	0.01		
	Total	24	5.70			

The findings indicate that all five models (MLR26–MLR30) are statistically significant at the 95% confidence level, as confirmed by the very low Significance F values (Significance F=0.00). Among these models, MLR26, MLR29, and MLR30 show the highest F values (174.83, 159.24, and 174.71, respectively), reflecting their strong explanatory power and consistency with the results reported in Table 12. MLR28 also demonstrates high statistical significance with an F value of 98.79, whereas MLR27, despite being significant, exhibits a comparatively lower F value (7.95), consistent with its weaker performance metrics. Overall, the ANOVA results reinforce that regression models integrating wind speed, temperature, and relative humidity are highly effective in estimating reference evapotranspiration.

3.5. Five predictor variables

Table 14 presents the regression coefficients and their statistical significance for the MLR model with five independent variables (MLR31).

Table 14. Summary of regression coefficients and statistical significance for an MLR model with five independent variables

Models	Variable	Coefficients	Standard Error	t Stat	P-value
MLR31	Intercept	1.26	0.97	1.30	0.21
	Minimum temperature	0.03	0.05	0.51	0.61
	Maximum temperature	0.08	0.05	1.44	0.17
	Relative humidity	-0.03	0.01	-3.79	0.00
	Wind speed	0.01	0.00	15.76	0.00
	Sunshine	0.04	0.09	0.50	0.62

The results show that among the predictors, relative humidity ($p < 0.01$) and wind speed ($p < 0.01$) have a statistically significant effect on ET_0 estimation. In contrast, minimum temperature, maximum temperature, and sunshine are not significant contributors, as their p-values are greater than 0.05. These findings suggest that, although including five variables improves the comprehensiveness of the model, not all predictors meaningfully influence evapotranspiration, and wind speed, together with relative humidity, remain the dominant factors. Table 15 summarizes the performance of the MLR model with five independent variables (MLR31).

Table 15. Performance of the regression model with five independent variables

Models	r	R^2	Adjusted R^2	SE	MAE	RMSE	NRMSE
MLR31	0.99	0.97	0.97	0.09	0.07	0.08	0.02

The MLR31 model demonstrates a very high correlation coefficient ($r = 0.99$), with a coefficient of determination of $R^2 = 0.97$ and an adjusted R^2 also equal to 0.97, indicating excellent explanatory power. The error indices, including SE (0.09), MAE (0.07), RMSE (0.08), and NRMSE (0.02), are all very low, confirming the strong predictive accuracy of the model in estimating reference evapotranspiration. Figure 6 presents a comparison between the ET_0 values estimated by the FAO-56 Penman-Monteith method and those predicted by the MLR model with five predictor variables (MLR31).

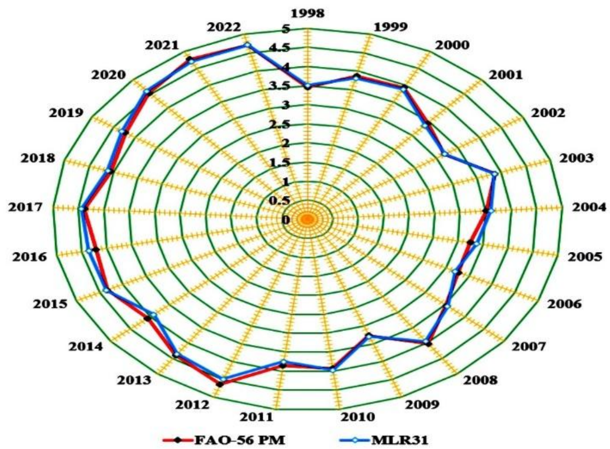

Fig. 6. Comparison of ET_0 Estimated by the FAO-56 Penman-Monteith Method and Regression Model with Five Predictor Variables

Figure 6 shows the close agreement between the two datasets, indicating the high accuracy and reliability of the MLR31 model, which is consistent with the performance metrics reported in Table 15. Table 16 summarizes the results of the analysis of variance for the MLR model with five independent variables (MLR31).

Table 16. ANOVA for the regression model with five independent variables

Models	Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Squares (MS)	F Value	Significance F
MLR31	Regression	5	5.54	1.11	134.68	0.00
	Residual	19	0.16	0.01		
	Total	24	5.70			

The regression source of variation shows a very high F-value (134.68) with a significance level of 0.00, indicating that the overall model is statistically significant. Out of the total sum of squares (5.70), the regression explains 5.54, while only 0.16 is attributed to residual error, confirming that the model accounts for nearly all the variability in reference evapotranspiration. These results are consistent with the high performance metrics reported in Table 15, further validating the reliability of MLR31. Finally, the five-variable model (MLR31) achieved $R^2 \approx 0.97$, matching the best-performing simpler models. These results emphasize that adding more variables does not necessarily improve predictive performance, and that carefully selected two- or three-variable models can provide estimation quality equivalent to more complex alternatives, while remaining more parsimonious and practical. The results of this study are consistent with those of Salahudin et al. (2023). The study by Salahudin et al. (2023) clearly supports the idea that well-selected smaller predictor sets can nearly match or exceed the performance of larger predictor sets in ET_0 modeling.

4. Conclusion

In this study, the evaluation of MLR models was conducted to estimate ET_0 at an annual scale in the DDR Basin of Iran. The main results of the study are as follows:

1) Single-variable models (MLR1–MLR5) revealed that minimum temperature, maximum temperature, and wind speed are statistically significant predictors of ET_0. Among them, wind speed alone (MLR4) demonstrated very high explanatory power ($R^2 \approx 0.92$), whereas other significant single-variable models explained a moderate portion of ET_0 variability ($R^2 \approx 0.39$).

2) Certain two-variable models, particularly MLR7, MLR12, MLR13, and MLR15, achieved very high predictive accuracy ($R^2 \approx 0.92–0.96$), comparable to more complex models with three or four predictors. This indicates that using just two carefully selected variables can provide near-optimal ET_0 estimation.

3) The three-variable models MLR16, MLR17, MLR19, MLR21, MLR23, and MLR25 achieved high accuracy, with coefficients of determination ranging from 0.93 to 0.97 and strong statistical significance, comparable to four- and five-variable models, whereas MLR18 and MLR22 showed weaker performance with R^2 below 0.55.

4) Four-variable models (MLR26–MLR30) generally achieved excellent predictive accuracy, with most models reaching $R^2 \approx 0.95–0.97$. However, MLR27 underperformed ($R^2 \approx 0.61$), suggesting that the inclusion of more variables does not automatically guarantee better model performance.

5) The five-variable model (MLR31) yielded $R^2 \approx 0.97$, matching the best performance observed in models with fewer predictors. This finding demonstrates that simpler models with two or three carefully selected predictors can be as effective as more complex formulations for ET_0 estimation.

This study is limited by its dataset, which consists solely of observations from the synoptic station in Kuhdasht, potentially restricting the applicability of the results to other regions or time periods. Moreover, the relatively small sample size (degrees of freedom ≈ 24) may introduce uncertainty in the estimated coefficients and limit the overall robustness of the regression models.

For future research, it is recommended to explore non-linear modeling approaches and artificial intelligence techniques, such as artificial neural networks or machine learning algorithms, to improve ET_0 prediction. Comparing these advanced methods with the current linear regression models may enhance accuracy and provide more robust tools for water resource management and agricultural planning.

Acknowledgments

Authors gratefully acknowledge the valuable feedback provided by the peer reviewers, which helped enhance the quality of this review.

Author Contributions

H. Ahmadpari: Data Collection, Data Analysis, Investigation, Methodology, Resources, Software, Writing – original draft, Writing – review and editing; **K. Vitaly:** Conceptualization, Validation, Supervision, Writing – review and editing; **Ata Amini:** Investigation, Methodology, Writing – original draft, Writing – review and editing.

Authors' Conflicts of interest

The authors declare no conflict of interest regarding the authorship or publication of this manuscript.

Data Availability Statement

All information and results are provided in the text of the article.

References

Ababakr, F. A., Ahmed, K. O., Amini, A., Karami Moghadam, M., & Gökçekuş, H. (2023). Spatio-temporal variations of groundwater quality index using geostatistical methods and GIS. Applied Water Science, 13(10), 206. doi: 10.1007/s13201-023-02010-4

Ahmadpari, H., & Khaustov, V. (2025a). Analyzing meteorological and hydrological droughts in the Darreh Dozdan River Basin through drought indices. Environment and Water Engineering, 11(2), 174-184.doi: 10.22034/ewe.2025.506959.2004

Ahmadpari, H., & Khaustov, V. (2025b). Agricultural drought monitoring using meteorological indices in Darreh Dozdan Basin, Iran. Advances in Civil Engineering and Environmental Science, 2(2), 72-84. doi: 10.22034/acees.2025.512324.1022

Ahmadpari, H., Hashemi Garmdareh, S. E., & Ghalehkohneh, K. (2017). Comparison of different methods of estimating potential evapotranspiration by FAO Penman Monteith (Case Study: Sepidan Region). Nivar, 41(98-99), 13-22. doi: 10.30467/nivar.2017.51886

Ahmadpari, H., Safavi Gerdini, M., & Ebrahimi, M. (2019). An appropriate method for estimating potential evapotranspiration in the absence of meteorological data (The case study of Khorrambid Township in Fars Province). Land Management Journal, 7(2), 223-230. doi: 10.22092/lmj.2019.120559

Ahmadpari, H., Shokoohi, E. S., Lalabadi, N. M., Gerdini, M. S., & Ebrahimi, M. (2019a). Assessment of potential evapotranspiration estimation methods in the fasa region. Specialty journal of agricultural sciences, 5(2-2019), 56-66.

Alazba, A. A., Mattar, M. A., El-Shafei, A., Radwan, F., Ezzeldin, M., & Alrdyan, N. (2025). Daily Reference Evapotranspiration Derived from Hourly Timestep Using Different Forms of

Penman–Monteith Model in Arid Climates. Water, 17(15), 2272. doi: 10.3390/w17152272

Amini, A. (2020). The role of climate parameters variation in the intensification of dust phenomenon. Natural Hazards, 102(1), 445–468. doi: 10.1007/s11069-020-03933-w

Amini, A., Abdeh Kolahchi, A., Al-Ansari, N., Moghadam, M. K., & Mohammad, T. (2019). Application of TRMM precipitation data to evaluate drought and its effects on water resources instability. Applied Sciences (Switzerland), 9(24), 5377. MDPI.

Amini, A., Arya, A., Eghbalzadeh, A., & Javan, M. (2017). Peak flood estimation under overtopping and piping conditions at Vahdat Dam, Kurdistan Iran. Arabian Journal of Geosciences, 10(6), 127. doi: 10.1007/s12517-017-2854-y

Amini, A., & Hesami, A. (2017). The role of land use change on the sustainability of groundwater resources in the eastern plains of Kurdistan, Iran. Environmental Monitoring and Assessment, 189(6), 297. doi: 10.1007/s10661-017-6014-3

Binesh, S., Ahmadpari, H., Shayegh, E., Masoumi, M., & Vakili Tajareh, F. (2020). Preparation of spatial distribution maps of saffron water requirement in Kermanshah province. International Journal of Engineering and Technology, 12(2), 321-336. doi: 10.21817/ijet/2020/v12i2/201202118

Brooks, G. P., & Ruengvirayudh, P. (2016). Best-subset selection criteria for multiple linear regression. General Linear Model Journal, 42(2), 14-25.

Cochran, W. G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. In Mathematical Proceedings of the Cambridge Philosophical Society (Vol. 30, No. 2, pp. 178-191). Cambridge University Press.

da Silva, H. J., dos Santos, M. S., Junior, J. B. C., & Spyrides, M. (2016). Modeling of reference evapotranspiration by multiple linear regression. Journal of Hyperspectral Remote Sensing, 6(1), 44-58. doi: 10.5935/2237-2202.20160005

Das, B. K., Jha, D. N., Sahu, S. K., Yadav, A. K., Raman, R. K., & Kartikeyan, M. (2022). Analysis of variance (ANOVA) and design of experiments. In Concept Building in Fisheries Data Analysis (pp. 119-136). Singapore: Springer Nature Singapore. doi: 10.1007/978-981-19-4411-6_7

Dehghani, T., Liaghat, A., Rezaei Rad, H., & Ahmadpari, H. (2024). Estimating grain corn yield based on Landsat 8 satellite images (Case study: agricultural and industrial lands of Shahid Beheshti, Dezful). Iranian Journal of Irrigation & Drainage, 18(3), 433-447.

Dimitriadou, S., & Nikolakopoulos, K. G. (2022). Multiple linear regression models with limited data for the prediction of reference evapotranspiration of the Peloponnese, Greece. Hydrology, 9(7), 124. doi: 10.3390/hydrology9070124

Djaman, K., Rudnick, D., Mel, V.C., Mutiibwa, D., Diop, L., Sall, M., Kabenge, I., Bodian, A., Tabari, H. and Irmak, S. (2017). Evaluation of Valiantzas' simplified forms of the FAO-56 Penman-Monteith reference evapotranspiration model in a humid climate. Journal of Irrigation and Drainage Engineering, 143(8), 06017005. doi: 10.1061/(ASCE)IR.1943-4774.0001191

Draper, N. R., & Smith, H. (1998). Applied regression analysis. McGraw-Hill. Inc. doi: 10.1002/9781118625590.ch15

Jiang, Y., & Sun, W. (2025). Day-ahead electricity price prediction and error correction method based on feature construction–singular spectrum analysis–long short-term memory. Energies, 18(4), 1-22. doi: 10.3390/en18040919

Karch, J. (2020). Improving on adjusted R-squared. Collabra: Psychology, 6(1), 45. doi: 10.1525/collabra.343

Karuppanan, S., Ramasamy, S., Lakshminarayanan, B., & Anuthaman, S. N. (2025). An effective machine learning model for the estimation of reference evapotranspiration under data-limited conditions. Research in Agricultural Engineering, 71(1), 22-37. doi: 10.17221/101/2023-RAE

Khadempour, F., Bakhtiari, B., & Golestani, S. (2017). Sensitivity Analysis of FAO Penman-Monteith Model in Daily Reference Evapotranspiration Estimation and Zoning Sensitivity Coefficients across Iran. Water and Soil, 31(4), 1046-1059. doi: 10.22067/jsw.v31i4.57602

Kim, H. Y. (2014). Analysis of variance (ANOVA) comparing means of more than two groups. Restorative dentistry & endodontics, 39(1), 74. doi: 10.5395/rde.2014.39.1.74

Koç, D. L., & Can, M. E. (2023). Reference evapotranspiration estimate with missing climatic data and multiple linear regression models. PeerJ, 11, e15252. doi: 10.7717/peerj.15252

Liu, D., Wang, Z., Wang, L., Chen, J., Li, C., & Shi, Y. (2024). Improved remote sensing reference evapotranspiration estimation using simple satellite data and machine learning. Science of the Total Environment, 947, 174480. doi: 10.1016/j.scitotenv.2024.174480

Miller, R. L., Acton, C., Fullerton, D. A., Maltby, J., & Campling, J. (2002). Analysis of variance (ANOVA). In SPSS for social scientists (pp. 145-154). London: Macmillan Education UK. doi: 10.1007/978-0-230-62968-4_8

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.

Purohit, R. C., Ingle, P. M., Bhakar, S. R., Mittal, H. K., Jain, H. K., Jain, H. K., & Singh, P. K. (2016). Impact of different meteorological parameters and relationship with short crop reference evapotranspiration for humid climatic conditions. International Research Journal of Earth Sciences, 4(8), 1-4.

Research Office of the Iran Water Resources Management Company. (2012). Guidelines and criteria for classification and coding of basins and study areas in Iran (No. 310). Department of Technical Affairs of the Ministry of Energy, Vice Presidency for Strategic Planning and Supervision.

Salahudin, H., Shoaib, M., Albano, R., Inam Baig, M.A., Hammad, M., Raza, A., Akhtar, A. and Ali, M.U. (2023). Using ensembles of machine learning techniques to predict reference evapotranspiration (et0) using limited meteorological data. Hydrology, 10(8), 169. doi: 10.3390/hydrology10080169

Stanley, R. P. (2011). Enumerative Combinatorics, Volume 1, Cambridge University Press, 725 pages.

Sureiman, O., & Mangera, C. M. (2020). F-test of overall significance in regression analysis simplified. Journal of the Practice of Cardiovascular Sciences, 6(2), 116-122. doi: 10.4103/jpcs.jpcs_18_20

Taheri, M., Bigdeli, M., Imanian, H., & Mohammadian, A. (2025). An Overview of Evapotranspiration Estimation Models Utilizing Artificial Intelligence. Water, 17(9), 1384. doi: 10.3390/w17091384

Tegos, A., Efstratiadis, A., Malamos, N., Mamassis, N., & Koutsoyiannis, D. (2015). Evaluation of a parametric approach for estimating potential evapotranspiration across different climates. Agriculture and Agricultural Science Procedia, 4, 2-9. doi: 10.1016/j.aaspro.2015.03.002

Usta, S., & Gençoğlan, S. (2019). Estimation of reference evapotranspiration using multiple linear regression models. International Journal of Scientific and Technological Research, 5(2), 5-19. doi: 10.7176/jstr/5-2-02

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2022). Probability and statistics for engineers and scientists. London, UK: Pearson.

Yirga, S. A. (2019). Modelling reference evapotranspiration for Megecha catchment by multiple linear regression. Modeling Earth Systems and Environment, 5(2), 471-477. doi: 10.1007/s40808-019-00574-2