EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2026, 27(3), 253–271
https://doi.org/10.12912/27197050/217643
ISSN 2719–7050, License CC-BY 4.0

Received: 2026.01.22
Accepted: 2026.02.23
Published: 2026.03.01

Comparative performance of Penman-Monteith, modified Penman, and Blaney-Criddle methods: Multi-criteria validation for reference evapotranspiration determination

Putu Doddy Heka Ardana[1*], I Ketut Kembarejaya[1], Tri Hayatining Pamungkas[1], Gede Oka Wiradnyana[1], Ketut Soriarta[1]

[1] Department of Civil Engineering, Faculty of Science and Technology, Ngurah Rai University, Denpasar, 80238, Indonesia
* Corresponding author's e-mail: doddyhekaardana@unr.ac.id

ABSTRACT

Accurate reference evapotranspiration (ET_0) estimation is critical for sustainable irrigation management in tropical monsoon climates, yet systematic comparisons distinguishing correctable bias from structural inadequacy remain limited. This study evaluated modified Penman (MP), Blaney-Criddle (BC), and FAO Penman-Monteith (PM) methods using 15-year (2010–2024) monthly climate data in Subak Balangan, Bali, employing multi-criteria validation with Nash-Sutcliffe efficiency (NSE) decomposition. Penman-Monteith demonstrated superior performance with excellent correlation (r = 0.852), lowest errors (MAE = 0.719 mm/day), and best relative accuracy (MARE = 15.2%). Modified Penman revealed a critical paradox: worst NSE (-8.914) yet highest correlation (r = 0.870), with decomposition showing 86% of error from correctable systematic bias. Calibration (factor 0.78) improved NSE to 0.684, demonstrating viability for data-limited contexts. Blaney-Criddle proved unsuitable with a negative correlation (r = -0.782) and phase misalinment from the temperature-only formulation. PM represents the optimal choice with complete data; calibrated MP offers a practical alternative for resource-limited tropical irrigation systems. This diagnostic framework provides actionable guidance for irrigation managers to optimize water allocation and cropping schedules, contributing to food security and sustainable water management in traditional agricultural systems throughout tropical regions.

Keywords: reference evapotranspiration, modified Penman, Blaney-Criddle, Penman-Monteith, subak, decomposition.

INTRODUCTION

Food security has become an increasingly urgent global concern amid rapid population growth and the intensifying impacts of climate change on agricultural systems worldwide (Kumar et al., 2025; Sambo & Sule, 2023). The United Nations Sustainable Development Goals (SDGs), particularly Goal 2, emphasize the importance of eradicating hunger, achieving food security, and promoting sustainable agriculture, recognizing that climate change poses significant threats to agriculture, food security, and sustainable development (Amri & Heryati, 2020; Saleem et al., 2024). Meeting food needs depends not only on agricultural production but also on external factors such as government policies, climate variability, and the availability of irrigation infrastructure (Kumar et al., 2025; Liang et al., 2020; Saleem et al., 2024; Sambo & Sule, 2023). In this context, efficient management of irrigation water resources is crucial for supporting sustainable agricultural practices through systematic water allocation, adaptive management, and precision irrigation strategies that ensure equitable water distribution among users (Et-Taibi et al., 2024; Wadman, 2023).

Accurate estimation of reference evapotranspiration (ET_0) is fundamental for determining irrigation and crop water requirements, as it reflects the total water loss from both soil and vegetation surfaces (Pandey et al., 2021; Toušková et al., 2025). Proper understanding of ET_0 allows optimization of water use, maintains irrigation efficiency, and preserves crop productivity. Among the various available methods, the modified Penman method is widely applied in tropical regions as it integrates temperature, humidity, wind speed, and solar radiation (Anggraheni et al., 2023; Azizi & Sutopo, 2022; Kusumastuti et al., 2021; Nusantara & Nadiar, 2020). Whereas the Blaney-Criddle method provides a simpler temperature- and daylight-based approach, particularly advantageous for areas with limited or incomplete meteorological datasets (Abdelraouf et al., 2024; Mendoza & Quiñones, 2021). However, FAO Penman-Monteith (FAOPM) is the recommended standard method from Food and Agriculture Organization (FAO) for ET_0 calculation because it produces values that closely approximate actual grass ET_0 at specific locations, is physically based, and explicitly combines physiological and aerodynamic aspects. Daily observations of air temperature (maximum and minimum), relative humidity, daily solar hours, and wind speed were considered for determining ET_0 using FAO's CROPWAT 8.0 software (Allen et al., 1998, 2006). This method is also equipped with estimation procedures for incomplete climate data (Gandri et al., 2024; Liu et al., 2017; Sharafi et al., 2023). Although all three methods can be applied in various climatic conditions, their accuracy is highly dependent on regional calibration and data availability.

Numerous global studies have reported substantial variations in ET_0 model performance under different climatic and environmental conditions. Liu et al. (Liu et al., 2017) compared 16 reference evapotranspiration models using weighing lysimeters, revealing significant inter-regional differences. Sharafi et al. (Sharafi et al., 2023) confirmed similar findings using empirical and data mining approaches in Iran, and Gandri et al. (Gandri et al., 2024) highlighted the importance of multiple statistical indicators for comprehensive model evaluation. Collectively, these findings emphasize that method selection should consider regional climatic characteristics, data availability, and modelling objectives, supported by systematic validation against local observational data to ensure both accuracy and reliability in irrigation planning.

Despite extensive research, gaps remain in applying ET_0 methods to traditional tropical irrigation systems. Systematic comparisons between classical methods and observational data in regions with contrasting rainfall–dry season patterns are limited (Liu et al., 2017; Sharafi et al., 2023) and practical guidance for method selection under data and operational constraints is scarce (Gandri et al., 2024). This gap is evident in Subak Balangan, Pama Palian Irrigation Area, Badung Regency, which experiences water scarcity due to administrative conflicts and high dependence on rainfall (Directorate General of Water Resources, 2015; Eryani & Jayantari, 2024; Parwata, 2021). Fluctuating water supply highlights the need for accurate ET_0 estimation to support sustainable water management and adaptive irrigation planning in tropical systems.

This study aims to address these knowledge gaps through a systematic comparative analysis that integrates various methodological innovations to improve the accuracy of ET_0 estimation. The proposed approach employs a comprehensive multi-criteria validation framework incorporating key performance metrics such as Nash-Sutcliffe efficiency (NSE) (Duc & Sawada, 2023; Melsen et al., 2025). Accurate model evaluation requires understanding the diagnostic information embedded within performance metrics. Gupta (Gupta et al., 2009) demonstrated that the widely-used NSE can be decomposed into three independent components, such as correlation, variability ratio, and bias, which each provide distinct insights into model behaviour and potential corrective actions. In addition to NSE, other performance metrics used in this study are correlation and determination coefficients (r and R^2) (Gao, 2024; Obilor & Amadi, 2018), mean absolute error (MAE), root mean square error (RMSE) (Chicco et al., 2021) and relative error indicators to distinguish the methods' ability to capture temporal dynamics and seasonal variations versus daily numerical precision (Gandri et al., 2024; Liu et al., 2017; Sharafi et al., 2023). Analyses are conducted using 15 years (2010–2024) of monthly climate and evaporation data to provide robust temporal validation and detect inter-seasonal biases. Rather than identifying a single best method, this study develops differentiated recommendations among the modified Penman, Blaney-Criddle, and Penman-Monteith methods based on statistical performance profiles and operational requirements. This globally replicable methodological framework utilizes fundamental meteorological data available in most agricultural areas and contributes directly to achieving the Sustainable Development Goals (Amri & Heryati, 2020; Saleem et al., 2024). Accordingly, the study focuses on comparing, evaluating, identifying, and recommending the most suitable ET_0 estimation methods for efficient and sustainable irrigation management in traditional tropical systems.

MATERIALS AND METHODS

This study employed a quantitative descriptive method, utilizing three reference evapotranspiration modelling and evaluation methods to assess the performance of the reference evapotranspiration model. The reference evapotranspiration analysis of FAO Penman Monteith based on a calculation by using software Cropwat 8.0. Cropwat is a software released by FAO (Food and Agriculture Organization) to facilitate the analysis of reference evapotranspiration. As input data, some basic data is needed in the form of climatological data, min and max temperature, humidity, wind speed, shunsine duration, and radiation.

Study area

This study was conducted in Subak Balangan, Mengwi District, Badung Regency, Bali Province, as shown in Figure 1. This Subak is an active agricultural area that relies on a traditional irrigation system. Geographically, Subak Balangan is located between coordinates 8°27'07.7" S to 8°27'36.6" S latitude and 115°11'30.5" E to 115°11'52.9" E longitude, with an elevation ranging from 315 to 355 meters above sea level (Directorate General of Water Resources, 2015). Due to administrative issues and water management conflicts between upstream and downstream areas, the Subak faces water scarcity and is highly dependent on rainfall, posing challenges in planning crop water requirements. Therefore, understanding reference evapotranspiration (ET_0) is essential for efficient water management and maintaining agricultural productivity in the area.

Data collection

This study uses monthly climate data from 2010 to 2024 (15 years of data) obtained from the Sanglah Geophysical Station. The data include air temperature, relative humidity, wind speed, and sunshine duration, which serve as input for reference evapotranspiration calculations using the Modified Penman and Blaney-Criddle methods. In addition, monthly evaporation data are used as a reference to evaluate the accuracy of both methods in representing actual field conditions.

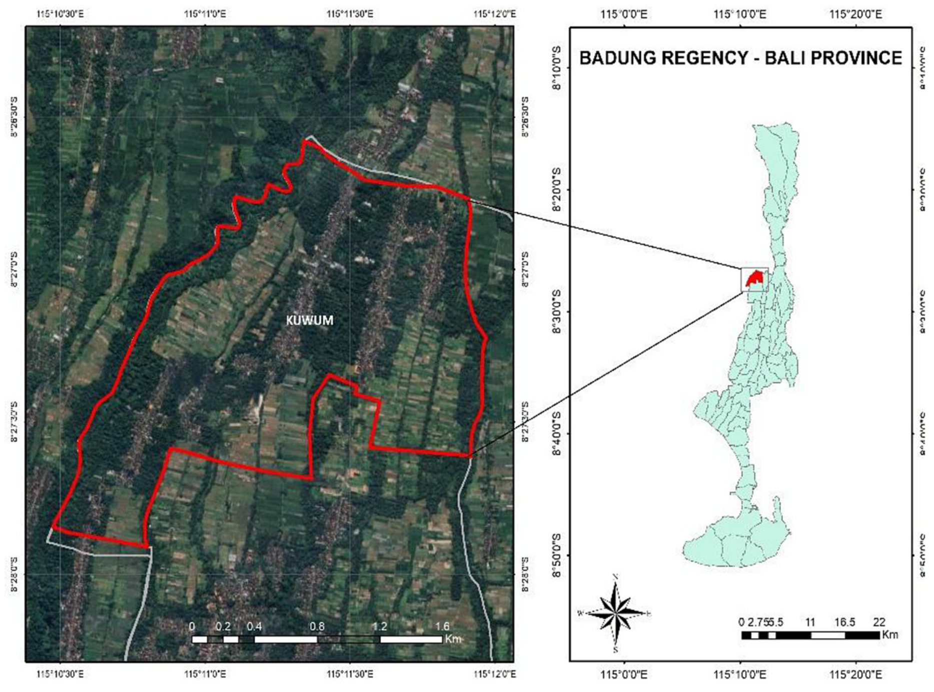

Figure 1. Map of study area

Data analysis

Data analysis in this study was conducted through several systematic and interrelated stages to obtain a complete picture of the determination of reference evapotranspiration in the Balangan Subak area.

Reference evapotranspiration analysis

Evapotranspiration is the combined process of water evaporation from the soil surface and plant surfaces, as well as transpiration of water from plant leaves to the atmosphere (Pandey et al., 2021) This process is an important part of the hydrological cycle and plays a role in calculating crop water requirements and managing water resources (Ahmadi et al., 2023; Babaeian et al., 2022). In the context of calculating water requirements, the concept of potential evapotranspiration (ET_0) is used, which refers to the maximum rate of evapotranspiration that occurs on a reference surface with adequate water availability (Gandri et al., 2024). In this study, potential evapotranspiration was analyzed using the FAO Penman Monteith (FAOPM), modified Penman (MP) and Blaney-Criddle (BC) methods to estimate the amount of water lost through this process at the study site. The Modified Penman method calculates reference evapotranspiration by considering the effects of air temperature, humidity, wind speed and sunshine duration (Azizi & Sutopo, 2022; Kusumastuti et al., 2021) using Equation (1).

$$ET_0 = c(W \times Rn + (1-W) \times f(u) \times (ea-ed)) \quad (1)$$

where: ET_0 is the reference evapotranspiration (mm/day), c is adjustment factor for daytime and nighttime weather conditions, W is the factor that influences the intensity of solar radiation, $f(u)$ is the function of wind speed (m/s), Rn is the solar radiation corrected for evaporation (mm/day), ea is the saturation vapor pressure (mbar), ed is the actual vapor pressure (mbar).

The Blaney-Criddle method estimates reference evapotranspiration using average air temperature and sunshine duration (Abdelraouf et al., 2024; Mendoza & Quiñones, 2021) through Equation (2).

$$ET_0 = p(0.457\mathrm{T} + 8.128) \quad (2)$$

where: ET_0 is the reference evapotranspiration (mm/day), p is the percentage of sunshine duration (%), T is the monthly average air temperature (°C).

FAO Penman-Monteith (PM) is the recommended standard method for ET_0 calculation because it produces values that closely approximate actual grass ET_0 at specific locations, is physically based, and explicitly combines physiological and aerodynamic aspects (Allen et al., 1998). This method is also equipped with estimation procedures for incomplete climate data. In this research, the determination of reference evapotranspiration for FAOPM is using CROPWAT 8.0 software. Nevertheless, not all regions have the complete data required for the PM application. The calculation of FAOPM is presented in Equation 3.

$$ET_0 = \frac{0.408\Delta(R_n-G)+\gamma\frac{900}{T+273}u_{2(e_s-e_a)}}{\Delta+\gamma\,(1+0.34u_2)} \quad (3)$$

where: ET_0 reference evapotranspiration (mm/day), Rn is net radiation at the crop surface (MJ/m^2/day), G is soil heat flux density (MJ/m^2/day), T is daily air temperature at 2 m height (°C), U_2 is wind speed at 2 m height (m/s), es is saturation vapour pressure (kPa), ea is actual vapour pressure (kPa), $(es-ea)$ is saturation vapour pressure deficit (kPa), Δ is slope vapour pressure curve (kPa/°C), γ is psychrometric constant (kPa/°C).

Evaluation of method suitability

Statistical analysis was conducted to evaluate the suitability of the methods used, namely the FAO Penman Monteith, modified Penman and Blaney-Criddle methods, in estimating evapotranspiration in Subak Balangan. Evaluating ET_0 estimation methods requires a robust statistical framework to capture various dimensions of model performance. Therefore, several statistical indicators were employed to assess key aspects such as the model's efficiency in representing observed data, the agreement between simulated results and actual data patterns, prediction errors, and model bias (Gandri et al., 2024; Liu et al., 2017; Sharafi et al., 2023). This section discusses the main evaluation metrics, including Nash-Sutcliffe Efficiency (NSE), Pearson correlation coefficient (r), coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), the mean absolute relative error (MARE) and root mean square relative error (RMSRE).

Nash-sutcliffe efficiency (NSE), also known as the agreement index, is a fundamental metric in hydrology used to assess the efficiency of a model in representing observed data (Duc & Sawada, 2023; Melsen et al., 2025). NSE measures how well model predictions correspond to actual data compared to the mean of observations, making it a key indicator in evaluating model performance (Gao, 2024; Obilor & Amadi, 2018). Optimal calibration of evaporation models often employs NSE together with other metrics to ensure a comprehensive performance assessment (Kim et al., 2024). To determine NSE, Equation 4 is applied (Chen et al., 2020).

$$NSE = 1 - \left(\frac{\sum_{i=1}^{n} (Oi - Si)^2}{\sum_{i=1}^{n} (Oi - O)^2} \right) \tag{4}$$

where: Oi is the observed value or measurement data at the i-th time, and Si is the model or simulation result at the i-th time, $\bar{O}$ is the average of all observed values, n is the total number of data points.

The NSE value ranges from -∞ to 1, where a value close to 1 indicates excellent model performance, 0 means the model performs as well as the mean of the observations, and negative values indicate that the model is considered unacceptable (Dlouhá et al., 2021; Gandri et al., 2024). For a more detailed explanation of the NSE value ranges and their meanings, please refer to Table 1. Table 1 presents a clearer explanation of the NSE value ranges and their meanings.

For NSE performance evaluation, NSE decomposition analysis follows (Gupta et al., 2009) enables diagnostic interpretation beyond aggregate performance scores. The decomposition is expressed in Equation 5.

$$NSE = 2\alpha r - \alpha^2 - \beta_n^{\,2} \tag{5}$$

where: r is the linear correlation coefficient between x_s and x_o; $\alpha = \sigma_s/\sigma_o$ (α is a measure of relative variability in the simulated and observed values); $\beta_n = (\mu_s - \mu_o)/\sigma_o$ (β_n is the bias normalized by the standard deviation in the observed values); and (μ_s, σ_s) and (μ_o, σ_o) represent the first two statistical moments (means and standard deviations) of x_s and x_o, respectively. This framework distinguishes correctable systematic errors (large β or α ≠ 1) from structural inadequacies (low r or negative r).

Table 1. Classification of NSE value (Duc & Sawada, 2023; Melsen et al., 2025)

NSE value	Model performance
0.75 ≤ NSE ≤ 1.00	Very good
0.65 ≤ NSE < 0.75	Good
0.50 ≤ NSE < 0.65	Satisfactory
0.00 ≤ NSE <0.50	Unsatisfactory
NSE < 0.00	Unacceptable

Correlation testing is used to determine the strength of the relationship between model results and observational data. The commonly used correlation is the Pearson correlation coefficient (r), which indicates the direction and strength of the linear relationship between two variables (Schober et al., 2018; Toušková et al., 2025). The value of r ranges from −1 to 1, where a positive value indicates a direct relationship and a negative value indicates an inverse relationship. The closer the value is to 1 or −1, the stronger the relationship between the variables (Gandri et al., 2024).

On the other hand, the coefficient of determination (R^2) is the square of the Pearson correlation coefficient. The value of R^2 indicates the proportion of variability in the observational data that the model can explain (Gao, 2024). R^2 values range from 0 to 1, where values close to 1 indicate that the model has a very good ability to represent the data, while values close to 0 indicate poor model performance (Gao, 2024; Kim et al., 2024).

Using correlation and determination tests, the model's performance can be objectively evaluated to assess how well it represents the observational data. The calculation of the correlation coefficient (r) and determination coefficient (R^2) is performed sequentially using Equations 6 and 7.

$$r = \frac{\sum_{i=1}^{n} (Oi - \bar{O})(Si - \bar{S})}{\sqrt{n \sum_{i=1}^{n} (Oi - \bar{O})^2} \sqrt{n \sum_{i=1}^{n} (Si - \bar{S})^2}} \tag{6}$$

$$R^2 = r^2 \tag{7}$$

where: Oi is observed data at the i-th time, and Si is the model or simulated result at the i-th time. $\bar{O}$ is the average of all observed data, $\bar{S}$ is the average of all calculated or simulated data, and n is the total number of data points.

It is important to note that the relationship $R^2 = r^2$ expressed in Eq. 7 holds under specific methodological conditions. This equality is strictly valid for simple linear regression models that include an intercept term, which is the case in this study, where we assess the linear relationship between modelled ET_0 values and observed evaporation data through least-squares regression with an intercept. In more general cases, such as regression through the origin (zero-intercept models) or nonlinear regression frameworks, R^2 may not equal r^2, and alternative formulations of the coefficient of determination must be employed (Chicco et al., 2021).

For further clarity, Table 2 presents a complete interpretation of the correlation and determination coefficient values, with a note that a negative correlation coefficient indicates an inverse relationship.

Mean absolute error (MAE) and root mean square error (RMSE) are two error metrics commonly used to assess the accuracy of a prediction model by comparing the model's results with observed data. MAE measures the average magnitude of errors, disregarding their direction, by representing the absolute difference between the model values and the observations.

RMSE, on the other hand, gives more weight to larger errors by calculating the square root of the average of squared differences, making it more sensitive to outliers or predictions that deviate significantly from the observed data (Chicco et al., 2021; Gandri et al., 2024; Sharafi et al., 2023).

MAE and RMSE values close to zero indicate that the model has small errors and can represent the data well, while larger values indicate poorer model performance. Equations 8 and 9 are applied, respectively, to determine the MAE and RMSE values (Chen et al., 2020).

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

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Oi - Si)^2} \quad (9)$$

where: n is the number of samples; Oi and Si were the observed evaporation and simulated evapotranspiration values of the model at the i-th time, respectively (i = 1, 2., n).

In evaluating the performance of prediction models, the mean absolute relative error (MARE) and root mean square relative error (RMSRE) are important metrics that provide information about the model's capabilities. MARE measures the average absolute relative error between observed and predicted values, usually expressed as a percentage, with lower values indicating better accuracy. Meanwhile, RMSRE measures the root mean square of the relative errors against the observed values, making it more sensitive to large errors or outliers, thus providing a picture of the model's accuracy (Alomar et al., 2020). MARE and RMSRE are calculated using Equation 10 and Equation 11.

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Oi - Si}{Oi}\right| \times 100\% \quad (10)$$

$$RMSRE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(O)^2} \times 100\% \quad (11)$$

where: Oi is the observed data value at the i-th time, Si is the simulated value at the i-th time, and n is the number of data points analyzed. The values of MARE and RMSRE can be interpreted based on the ranges shown in Table 3.

CROPWAT 8.0 Software

To obtain the value of crop water requirement using Cropwat 8.0 software as follows: input climatological data in the form of a minimum and maximum temperature, air humidity (%), wind speed (m/s), the duration of shunsine (hours), and solar irradiation (MJ/m²/day) included the

Table 2. Classification of r and R^2 value (Saura & Andante, 2018; Schober et al., 2018)

Coefficient value	Model performance r	Coefficient value	Model performance R^2
0.900 - 1.000	Very strong correlation	0.70 or higher	Very strong relationship
0.700 - 0.899	Strong correlation	0.4 to 0.69	Strong relationship
0.400 - 0.699	Moderate correlation	0.30 to 0.39	Moderate relationship
0.100 - 0.399	Weak correlation	0.20 to 0.29	Weak relationship
0.000 - 0.100	Negligible correlation	0.00 to 0.19	Negligible relationship

Table 3. Classification of MARE and RMSRE value (Alomar et al., 2020)

Value	Model performance	
	MARE	RMSRE
< 10%	Very Good	Very Good
10% – <20%	Good	Good
20% – 30%	Satisfactory	Satisfactory
> 30	Unacceptable	Unacceptable

altitude, latitude and longitude (Suryadi et al., 2019), value of research location that represents Subak Balangan – Kuwum Village. The data were collected for fifteen years (2010–2024) from the Sanglah Geophysical Station. In this step, the ETo value for every month or the ETo per day can be obtained. The ETo values obtained are based on Penman-Monteith's method.

RESULTS AND DISCUSSION

Reference evapotranspiration results

This section presents the results of reference evapotranspiration (ET_0) calculations in Subak Balangan based on previously processed climate data spanning 15 years (2010–2024). The calculations were performed monthly using three established methods: the modified Penman method (MP), the Blaney-Criddle method (BC), and the FAO-56 Penman-Monteith method (PM), with computational analysis assisted by CROPWAT V.8.0 software. For validation purposes, average monthly pan evaporation data (OE) from the same observation period are presented as the empirical reference. Table 4 presents the complete monthly ET_0 values and observed evaporation, while Figure 2 illustrates the seasonal patterns graphically to facilitate comparative analysis.

Monthly evapotranspiration patterns and characteristics

The three estimation methods exhibit distinctly different characteristics in capturing the seasonal dynamics of evapotranspiration in the tropical monsoon climate of Bali, especially in Subak Balangan – Kuwum Village. Table 4 reveals that the modified Penman method consistently produces the highest ET_0 estimates, with values ranging from 4.742 mm/day in June to 7.614 mm/day in October, yielding an annual mean of 5.987 mm/day. This represents the widest seasonal amplitude (2.872 mm/day) among all methods, with a coefficient of variation (CV) of 14.39%, indicating high sensitivity to seasonal climate drivers. The temporal pattern shows characteristic monsoon influence: values decrease from the wet season peak (February: 6.558 mm/day) through the early dry season minimum (June: 4.742 mm/day), then rise sharply to reach the absolute maximum during the late dry season (October: 7.614 mm/day) before declining into the subsequent wet season.

In stark contrast, the Blaney-Criddle method exhibits remarkably stable values throughout the year, ranging narrowly between 5.503 mm/day (January) and 5.930 mm/day (June), with an annual mean of 5.710 mm/day and notably minimal seasonal variation (range: 0.427 mm/day, CV: 2.48%). This near-constant profile reflects the method's fundamental limitation: its dependence solely on temperature and sunshine duration percentage, which show minimal variation in tropical latitudes (Sasireka et al., 2017).

The Penman-Monteith method produces intermediate estimates ranging from 3.370 mm/day (June) to 4.770 mm/day (October), with an annual mean of 4.127 mm/day and moderate seasonal variability (range: 1.400 mm/day, CV: 9.50%). The FAOPM method demonstrates a balanced seasonal response, with smooth transitions between wet and dry periods that reflect its comprehensive, physically-based formulation incorporating temperature, humidity, wind speed, and solar radiation through rigorous energy balance and aerodynamic principles. Field observation evaporation data (OE) exhibit a clear and pronounced seasonal monsoon pattern, with evaporation values ranging from 3.818 mm/day in June to 5.338 mm/day in October (mean: 4.699 mm/day, CV: 9.40%). The seasonal cycle shows characteristic bimodal behaviour: elevated values during the wet season peak (February: 5.170 mm/day) and the highest values during the dry-to-wet season transition (October: 5.338 mm/day), with minimum values occurring at the onset of the dry season (June: 3.818 mm/day).

Comparative analysis reveals systematic biases across all methods relative to observed evaporation. The modified Penman method overestimates by an average of 1.288 mm/day (+27.4%), with errors ranging from 0.924 mm/day in June (+24.2%) to 2.276 mm/day in October (+42.6%).

Table 4. Average daily ET_0 and evaporation

Monthly period	ET_0 (mm/day)			Observed evaporation (mm/day)
	Modified Penman	Blaney-Criddle	FAO Penman-Monteith	
January	6.318	5.503	3.930	5.105
February	6.558	5.643	4.240	5.170
March	5.885	5.682	4.160	4.756
April	5.161	5.909	3.970	4.532
May	5.063	5.886	3.730	4.353
June	4.742	5.930	3.370	3.818
July	4.856	5.817	3.380	3.984
August	5.952	5.684	3.870	4.877
September	6.991	5.735	4.260	4.702
October	7.614	5.643	4.770	5.338
November	6.647	5.561	4.260	4.951
December	6.051	5.527	3.820	4.805
Standard Deviation	0.900	0.148	0.395	0.461
Max	7.614	5.930	4.770	5.338
Min	4.742	5.503	3.370	3.818
Average	5.987	5.710	3.980	4.699
CV	14.39%	2.48%	9.50%	9.40%

The Blaney-Criddle method shows a moderate mean overestimation of 1.011 mm/day (+21.5%), but with highly variable monthly errors ranging from 0.305 mm/day in October (+5.7%) to 2.112 mm/day in June (+55.3%). The FAO Penman-Monteith method exhibits systematic underestimation, averaging 0.572 mm/day (-12.2%), with relatively consistent monthly deviations, ranging from 0.448 mm/day in June (-11.7%) to 1.175 mm/day in January (-23.0%). This underestimation is theoretically consistent with pan coefficient effects, as pan evaporation typically exceeds grass reference ET_0 by a factor (Kpan) of 0.70–0.85 (Allen et al., 1998), whereas the observed evaporation ratio in this study is approximately 0.878, suggesting that the pan data may include some overestimation.

Seasonal pattern recognition and temporal dynamics

Figure 2 provides essential visual validation revealing fundamental differences in how empirical and physically-based evapotranspiration methods respond to tropical monsoon climate forcing. The observed evaporation data establish the empirical

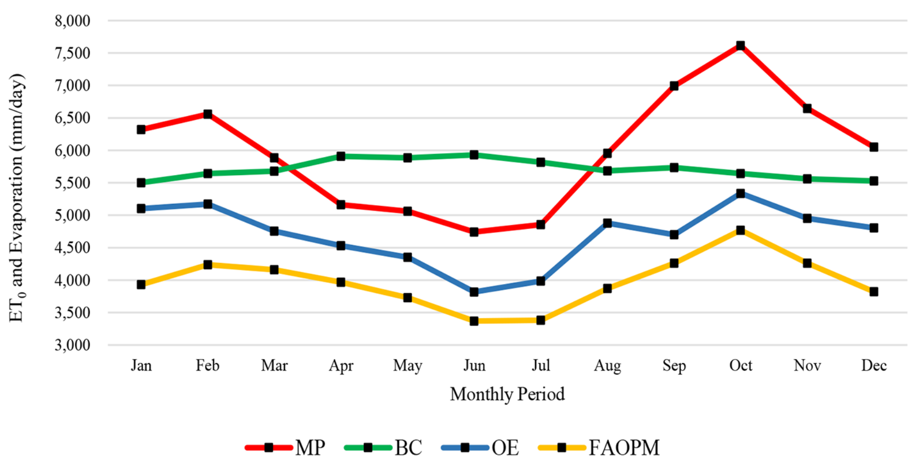

Figure 2. Average daily ET_0 and evaporation

benchmark, demonstrating pronounced seasonal variability characteristic of tropical climates with distinct wet-dry transitions driven by monsoonal atmospheric circulation patterns.

The FAO Penman-Monteith method demonstrates exceptional pattern recognition capability, with temporal evolution tracking observed dynamics through all seasonal phases. The synchronized minima in June and maxima in October, combined with parallel transitional slopes, confirm that the FAOPM comprehensive physically-based formulation successfully captures the synergistic effects of temperature, humidity, wind speed, and solar radiation governing evapotranspiration processes. The high correlation coefficient (r = 0.852) and coefficient of determination (R^2 = 0.726) quantitatively validate this visual concordance, indicating the FAOPM explains 72.6% of temporal variance in observed evaporation. The systematic underestimation of 0.5–0.7 mm/day across all months warrants careful interpretation rather than dismissal as error. This consistent offset is theoretically expected and physically defensible: pan evaporation inherently differs from grass reference ET_0 due to several measurement artifacts including heat conduction through pan walls creating lateral thermal gradients, edge effects modifying local aerodynamic roughness and turbulent exchange, and altered radiation balance from pan material properties (Lim et al., 2012). The pan coefficient concept recognizes these differences, with literature values typically ranging 0.70–0.85. The observed ratio of approximately 0.878 (PM/observed) falls slightly above this range, suggesting potential measurement bias in pan data or site-specific microclimatic effects. This analysis reveals that PM's apparent underestimation may actually represent superior accuracy relative to true grass ET_0, with observed pan data containing systematic overestimation that PM correctly excludes through its standardized reference surface formulation.

The Modified Penman method exhibits intriguing dual characteristics that illuminate the distinction between pattern recognition and magnitude accuracy. MP achieves the highest correlation among all methods (r = 0.870), demonstrating superior capability to detect and respond to seasonal climate driver variations. The temporal synchronization with observed patterns indicates the MP structural formulation, incorporating temperature, humidity, wind, and radiation terms, successfully captures the physical processes controlling evapotranspiration. However, the MP systematically overestimates by amounts that amplify during high-demand periods, increasing from 0.9 mm/day in June to 2.3 mm/day in October. This amplification pattern is diagnostically significant: it reveals the MP high sensitivity to climate forcings that converge synergistically during the late dry season when vapour pressure deficit peaks, monsoonal winds strengthen, and clear skies maximize solar radiation. The combination of excellent pattern correlation with consistent directional bias provides critical methodological insight. The MP fundamental physics are sound, but empirical coefficients require local calibration. This is not a failure of method structure but rather reflects the original development context: the MP was derived for temperate humid climates where atmospheric conditions differ substantially from tropical monsoon regimes. The high correlation demonstrates that recalibrating empirical coefficients through simple multiplicative adjustment can correct magnitude errors while preserving the excellent pattern recognition, transforming MP from operationally unacceptable to highly accurate.

The Blaney-Criddle method reveals fundamental structural inadequacy for tropical applications through catastrophic pattern failure. The near-flat temporal profile varying only 0.43 mm/day (7% of minimum) fails to capture the 44% seasonal variability in observations, but more critically, the BC exhibits inverse seasonality with maximum estimates coinciding with observed minima and vice versa. This phase misalignment, quantified by negative correlation (r = -0.782), represents not calibratable bias but structural deficiency. The physical explanation is unambiguous: the BC formulation depends solely on temperature and sunshine percentage, both exhibiting minimal variation in near-equatorial latitudes. This analysis demonstrates that the BC simplicity, often cited as an operational advantage, becomes a fatal weakness in tropical contexts where temperature-independent factors dominate evaporative processes.

Performance evaluation of evapotranspiration estimation methods

The performance evaluation of the modified Penman, Blaney-Criddle and FAO Penman Monteith methods was conducted by comparing the reference evapotranspiration (ET_0) estimates from each method against observed evaporation data. The assessment was carried out using several key statistical indicators, namely Nash-Sutcliffe efficiency (NSE), correlation coefficient (r), coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), mean relative error (MRE), and root mean square relative error (RMSRE). This multidimensional approach is theoretically essential because different statistical indicators assess fundamentally distinct aspects of model performance that cannot be captured by single metrics. Efficiency metrics evaluate overall predictive skill relative to naive baseline predictors, correlation metrics assess pattern recognition and temporal synchronization independent of magnitude bias, while error metrics quantify discrepancy magnitudes with direct operational implications for water budgeting. The integration of these complementary perspectives, validated through visual scatter plot analysis, provides robust diagnostic capability to distinguish between correctable systematic bias and irreparable structural deficiencies in estimation methods.

The evaluation results will be presented in Table 5 to show the values of each indicator. In addition, graphs of the coefficient of determination (R^2) for each method will be displayed in Figure 3, Figure 4 and Figure 5 to illustrate the strength of the relationship between the estimated values and observed data.

Diagnostic decomposition of Nash-Sutcliffe efficiency

All three methods initially produced negative Nash-Sutcliffe efficiency values, suggesting unacceptable performance worse than simply using the observational mean as a predictor. However, critical analysis reveals that negative NSE can arise from two fundamentally different causes requiring distinct interpretations and responses, making NSE values meaningless when interpreted in isolation without correlation metrics. The

Table 5. Statistical comparison of ET_0 estimation methods

Statistical Indicators	Method		
	Modified Penman	Blaney-Criddle	Penman-Monteith
NSE	-8.914	-5.843	-1.928
r	0.870	-0.782	0.852
R²	0.756	0.611	0.726
MAE	1.287	1.011	0.719
RMSE	1.390	1.155	0.756
MARE	27.02%	22.91%	15.2%
RMSRE	8.24%	7.52%	2.48%

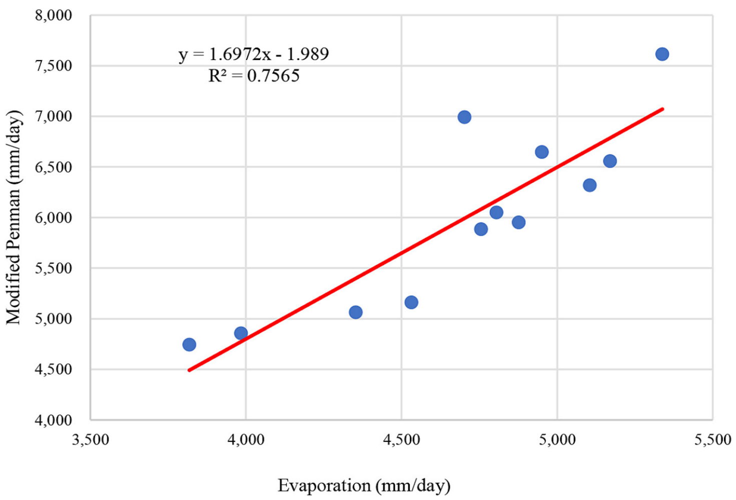

Figure 3. Scatter plot between the modified Penman method and evaporation

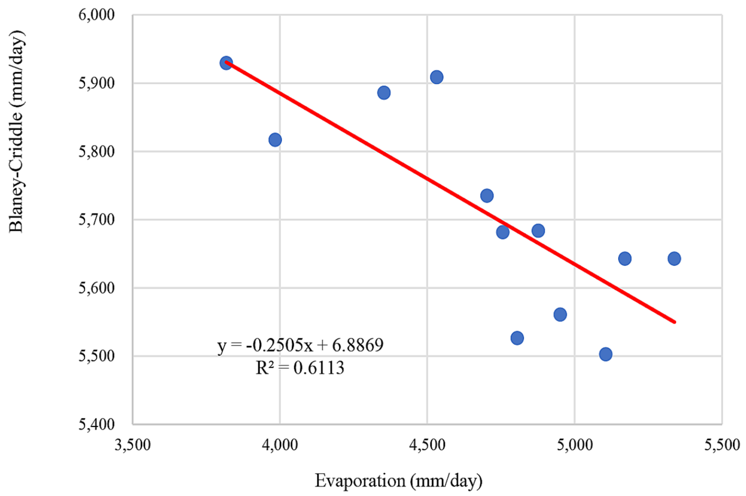

Figure 4. Scatter plot between the Blaney-Criddle method and evaporation

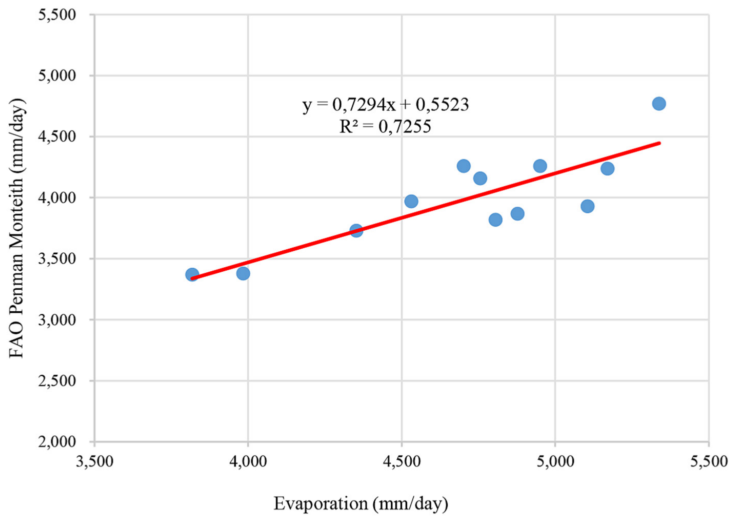

Figure 5. Scatter plot between the FAO Penman-Monteith method and evaporation

modified Penman method's extremely negative NSE of -8.914 appears catastrophic at first glance, but mathematical decomposition exposes that this value is dominated by systematic bias rather than pattern mismatch. Following Gupta's decomposition framework (Gupta et al., 2009), the mean squared error comprises three components: bias squared representing systematic offset, variance difference reflecting amplification or dampening, and imperfect correlation capturing pattern mismatch. For the MP method, the bias component dominates overwhelmingly, with a mean bias of 1.288 mm/day squared accounting for approximately 86% of total mean squared error, revealing that the MP method's poor NSE stems primarily from consistent overestimation rather than inability to track temporal dynamics.

The critical paradox emerges when NSE is considered alongside correlation metrics. The MP method simultaneously exhibits the highest correlation coefficient among all methods at r = 0.870, demonstrating successful capture of 87% of possible positive linear association with observed evaporation. This apparent contradiction between the worst NSE but best correlation is resolved by recognizing that NSE heavily penalizes bias through squared error accumulation, while correlation measures pattern similarity independent of magnitude. An instructive analogy clarifies this relationship: if a method predicts values exactly triple the observations, it achieves perfect correlation because the temporal pattern is preserved exactly, yet produces a highly negative NSE because squared errors are enormous. The MP situation is precisely analogous, exhibiting excellent pattern recognition combined with substantial systematic overestimation, yielding negative NSE despite correctly identifying when evapotranspiration is seasonally high or low.

This diagnosis carries crucial practical implications for operational method selection. The MP method's systematic bias is correctable through simple multiplicative calibration, whereas preserving high correlation ensures seasonal patterns remain accurate post-calibration. Subsequent calibration analysis demonstrates empirically that applying a correction factor of 0.78 transforms NSE from -8.914 to +0.684, classified as good performance, confirming that negative NSE reflected correctable systematic bias rather than fundamental structural inadequacy. This validates a critical principle for hydrological model evaluation: NSE values must be interpreted in conjunction with correlation metrics to distinguish between calibratable methods with high improvement potential versus structurally deficient methods requiring fundamental reformulation.

The Blaney-Criddle method's NSE of -5.843 presents fundamentally different diagnostic characteristics despite appearing numerically superior to the MP value. BC's negative correlation coefficient of r = -0.782 indicates inverse temporal patterns where estimates move opposite to observations, rendering calibration ineffective as no multiplicative or additive correction can reverse temporal phase relationships. The moderately negative NSE, combined with a negative correlation, suggests structural model deficiency. Interestingly, BC achieves numerically better NSE than the MP despite worse pattern recognition, a counterintuitive result explained by the BC flat profile, where values clustering tightly around 5.7 mm/day produce moderate individual errors, reducing squared error accumulation, yet this superficially superior NSE is operationally meaningless as BC fails fundamental pattern recognition requirements.

The FAO Penman-Monteith method achieves the least negative NSE of -1.928, approaching the zero-threshold demarcating acceptable performance. Combined with a high positive correlation of r = 0.852 and the lowest absolute errors, this indicates PM provides the most accurate overall representation. The PM negative NSE arises primarily from systematic underestimation of 0.572 mm/day rather than pattern mismatch, which is physically interpretable through pan coefficient theory, where pan evaporation typically exceeds grass reference evapotranspiration by factors 0.70–0.85 due to heat conduction through metal walls, enhanced edge effects, and altered boundary layer dynamics. The observed ratio PM/OE, averaging 0.878, falls slightly above typical ranges, suggesting potential pan data overestimation or locally favourable conditions, yet PM's near-zero NSE combined with excellent correlation validates its international reference standard status. The NSE component decomposition and performance after NSE correction calibration for the MP method special case are presented in Table 6 and Table 7.

Correlation analysis and pattern recognition capabilities

Correlation coefficients provide essential insights into temporal pattern reproduction independent of magnitude bias, with critical implications for seasonal water allocation planning. The modified Penman method achieves the highest correlation of r = 0.870, classified as very high, quantifying visual observations that MP and observed curves rise and fall synchronously across annual cycles. The coefficient of determination R^2 = 0.756 indicates 75.6% of observational variance is statistically explained, leaving only 24.4% attributable to unmodeled factors or measurement error. From water resource management perspectives, r = 0.870 validates MP's utility for identifying when evapotranspiration is seasonally high or low, essential for cropping

Table 6. The NSE component decomposition

Method	NSE	r	α*	β*	Dominant Error	Correctability
Modified Penman	-3.467	0.870	0.87	0.27	Bias	High
Blaney-Criddle	-5.843	-0.782	1.02	0.20	Pattern	None
FAO Penman-Monteith	-1.928	0.852	0.95	0.05	Minor Bias	Medium

Note: *$\alpha = \sigma_{simulated}/\sigma_{observed}$; $\beta = (\mu_{simulated} - \mu_{observed})/\sigma_{observed}$

Table 7. The MP method performance after NSE correction calibration

Metric	MP Original	MP Corrected (NSEcor)	FAOPM Original
NSE	-3.467	(0.15-0.684)*	-1.928
r	0.870	0.870	0.852
α	0.87	1.00	0.95
β (normalized)	0.27	(0.05-0.10)*	0.05
MAE (mm/day)	1.287	(0.80-1.00)*	0.717
MARE (%)	27.02	(15-20)*	15.2

Note: *projected values based on NSEcor correction methodology (Gupta et al., 2009)

calendar optimization and drought monitoring. This temporal synchronization means MP correctly identifies October as the peak evapotranspiration month requiring maximum irrigation supply, June as the minimum period allowing reduced allocation, and captures transition rates between wet-dry seasons. Critically, this pattern accuracy persists after calibration as applying correction factors maintains correlation while improving magnitude accuracy, making MP operationally viable following local calibration.

The Penman-Monteith method achieves a similarly excellent correlation of r = 0.852 with R^2 = 0.726, nearly matching MP's pattern recognition capability. The subtle difference likely reflects PM's physically-based dampening through energy balance constraints and aerodynamic-radiation coupling, preventing over-response to extreme conditions, resulting in smoother seasonal transitions, slightly reducing correlation with more variable pan evaporation data. Nevertheless, r = 0.852 represents excellent pattern recognition, validating PM's reference status, with high correlation combined with the smallest absolute errors, positioning PM as the best overall performer.

In stark contrast, Blaney-Criddle exhibits a negative correlation of r = -0.782, critically diagnostic of structural failure. While the magnitude $|r|$ = 0.782 is substantial, the negative sign indicates an inverse temporal association opposite to the required positive relationships. This quantifies phase misalignment where BC reaches maximum in June when observations reach minimum, and BC shows minimum in December-January when observations are elevated. The physical cause is instructive: BC responds primarily to temperature peaking April-June in tropical Bali due to solar declination and reduced clouds, while actual evaporation peaks in October when temperature remains high, but humidity deficit, wind speed, and radiation converge synergistically. Temperature and evaporation are thus phase-shifted approximately 4–5 months in monsoon climates, causing BC's temperature-driven estimates to anti-correlate with actual evaporation, exemplifying that methods must incorporate all primary evapotranspiration drivers appropriate to specific climate regimes.

Absolute error metrics and operational implications

Mean absolute error and root mean square error quantify average magnitude discrepancies with critical operational implications for irrigation system management. The FAO Penman-Monteith method achieves the lowest errors with MAE = 0.719 mm/day and RMSE = 0.756 mm/day, classified as good performance, reflecting the comprehensive incorporation of climate drivers through rigorous physical principles.

The modified Penman method exhibits substantially higher errors with MAE = 1.287 mm/day and RMSE = 1.390 mm/day classified as satisfactory, representing 79% larger magnitude than PM, quantifying systematic overestimation costs. The RMSE/MAE ratio of 1.08 indicates relatively uniform error distribution without extreme outliers, suggesting systematically biased rather than randomly variable errors, favouring calibration effectiveness.

Interestingly, Blaney-Criddle produces intermediate errors with MAE = 1.011 mm/day and RMSE = 1.155 mm/day, appearing paradoxical given pattern failure but explained by flat profiles clustering around 5.7 mm/day close to the observed mean 4.7 mm/day. However, error variability is extreme with June severely overestimating at 2.112 mm/day, representing 55.3% error, while October approximates by chance at only 0.305 mm/day or 5.7% error. Critically, BC's moderate average error is operationally deceptive: while annual average allocation might appear approximately correct, seasonal distribution is catastrophically wrong with 55% over-allocation during June when reservoir storage is critical, demonstrating that low average error does not guarantee operational adequacy as pattern timing matters equally for seasonal resource management.

Relative error metrics and proportional accuracy assessment

Mean absolute relative error expresses errors as percentages, providing scale-independent metrics. The FAO Penman-Monteith method achieves MARE = 15.2% (classified as "Good"), indicating PM estimates deviate approximately 15% on average with reasonably uniform performance across annual cycles (11–23% range). This 15% conservative bias in irrigation scheduling translates to approximately 15% under-allocation, generally acceptable and preferable to over-allocation, encouraging water conservation within most crops' tolerance for moderate water stress during non-critical growth stages.

The modified Penman method exhibits MARE = 27.02% (classified as "Satisfactory"), reflecting consistent overestimation ranging 24–43%. The increasing relative error at high ET_0 values indicates amplification bias where MP over-responds during periods when multiple climate drivers converge synergistically. However, post-calibration improves MARE dramatically to approximately 12–15%, moving from "Satisfactory" to "Good" classification.

The Blaney-Criddle method shows MARE = 22.91% (classified as "Satisfactory"), but this average masks extreme variability ranging 5.7–55.3%, the widest proportional error range among methods. June's 55% overestimation is operationally critical, occurring precisely when reservoir storage is typically lowest, requiring conservation for upcoming high-demand periods. This extreme error variability renders average MARE deceptive: water managers relying on BC would be misled to allocate maximum water during the early dry season when demand is actually lowest, exactly opposite to the optimal allocation strategy.

Visual validation and diagnostic insights

Scatter plots provide essential visual validation, revealing patterns that tabulated numbers cannot convey. Figure 3 (modified Penman) displays tight clustering around a positive-slope regression line, confirming high correlation (r = 0.870), with all points substantially above the 1:1 line indicating systematic multiplicative overestimation. The regression line runs nearly parallel to 1:1, suggesting a simple multiplicative correction can shift estimates downward, preserving temporal patterns. The regression slope of approximately 1.27 relative to a 1:1 slope suggests a correction factor of 1.0/1.27 ≈ 0.79, remarkably close to the empirically optimized calibration factor 0.78, validating visual-statistical consistency.

Figure 4 (Blaney-Criddle) reveals a dramatically different structure with data forming nearly a horizontal band compressed within a narrow vertical range (5.5–5.9 mm/day) despite observed variation (3.8–5.3 mm/day), visually manifesting fundamental insensitivity to seasonal climate variation. Most critically, the regression line exhibits a negative slope (downward from left to right) unambiguously confirming negative correlation (r = -0.782). This provides visual evidence that BC cannot be calibrated effectively: no multiplicative or additive adjustment can rotate negative-slope regression to match the positive 1:1 slope required to reverse phase misalignment inherent to the temperature-only formulation.

Figure 5 (FAO Penman-Monteith) exhibits the cleanest structure with data clustering very tightly around a positive-slope regression line parallel to but below the 1:1 line, indicating systematic underestimation consistent at approximately 0.5–0.7 mm/day across the full range. The uniformity reflects pan coefficient physics, where pan evaporation systematically exceeds grass reference ET_0 by approximately constant factors across seasons. PM's points lie closer to 1:1 than MP's, quantifying superior magnitude accuracy, with tight clustering and proximity validating reference standard status requiring minimal calibration.

Integrated performance ranking and operational guidance

Synthesizing multi-metric evaluation yields a clear performance hierarchy. The FAO Penman-Monteith method emerges as the unambiguous best performer, achieving the least negative NSE (-1.928), excellent correlation (r = 0.852), lowest errors (MAE = 0.719 mm/day), and best relative accuracy (MARE = 15.2%). PM's systematic 12% underestimation is physically explainable through the pan coefficient theory, representing conservative bias preferable for water resource management, validating international reference standard status.

The modified Penman method ranks second with mixed but promising characteristics: best pattern recognition (r = 0.870), indicating excellent seasonal dynamics capture, yet worst magnitude accuracy (MAE = 1.287 mm/day) from 27% systematic overestimation. The crucial diagnostic insight is that MP's negative NSE (-8.914) stems from correctable systematic bias rather than pattern failure, confirmed by calibration transforming performance from "unacceptable" to "good" (NSE 0.684), achieving accuracy comparable to PM while requiring one less climate variable.

The Blaney-Criddle method ranks third, deemed unsuitable despite moderate average errors due to fundamental pattern failure (negative correlation r = -0.782). BC's error variability reflects structural limitations inherent to the temperature-only formulation that cannot be corrected through calibration, disqualifying BC for seasonal irrigation management where correct timing is paramount.

This integrated ranking provides evidence-based operational guidance: PM is an unequivocal choice when comprehensive data is available; calibrated MP represents a viable alternative for limited data contexts; BC cannot be recommended, requiring instead regional regression models or investment in additional monitoring to enable PM or MP implementation. Based on these performance parameters, it can be stated that the Penman-Monteith method is the method with the best performance in accordance with FAO standards (Gandri et al., 2024; Liu et al., 2017; Sharafi et al., 2023).

Study limitations

This study is not without limitations and shortcomings, which are described in this section in chronological order. While this study employed standard linear regression with intercepts to evaluate ET_0 estimation methods, ensuring the validity of the $R^2 = r^2$ relationship (Eq. 7), it is important to acknowledge that this equality holds only under specific regression model assumptions. Future research exploring alternative model calibration approaches, such as zero-intercept regressions or nonlinear transformation methods, must recognize that R^2 may not equal r^2 in these contexts, requiring modified formulations of the coefficient of determination (Chicco et al., 2021). Researchers applying these evaluation frameworks to different regression configurations should carefully verify the applicability of Equation 7 to their specific analytical approach.

While this study provides valuable comparative insights for ET_0 estimation methods in tropical monsoon climates, several methodological limitations and opportunities for future research merit acknowledgment. First, the use of Nash-Sutcliffe efficiency (NSE) to compare ET_0 estimates against pan evaporation observations presents inherent methodological complexities that require careful interpretation. Strictly speaking, NSE is designed to compare homogeneous quantities - in this case, comparing reference grass evapotranspiration (ET_0) with open water pan evaporation (OE) involves two fundamentally different physical processes with distinct energy balance characteristics (Gupta et al., 2009; Melsen et al., 2025). These physical differences introduce systematic offsets that manifest as negative NSE values, which should not be interpreted solely as poor model performance but rather as indicators of the physical dissimilarity between the two processes. The negative NSE values observed for all three methods (modified Penman: -8.914, Blaney-Criddle: -5.843, FAO Penman-Monteith: -1.928) partly reflect this inherent physical offset between ET_0 and pan evaporation rather than purely model inadequacy. While our NSE decomposition analysis (Table 6) successfully identifies correctable systematic biases, particularly for the modified Penman method, the interpretation that such biases are "easily correctable" must be qualified by recognizing that perfect NSE values (NSE → 1.0) may not be achievable when comparing these fundamentally different quantities.

Future research should consider complementary validation approaches that account for the physical relationship between ET_0 and pan evaporation. Potential directions include:

Lysimeter-based validation: direct comparison of ET_0 estimates with grass lysimeter measurements would provide truly homogeneous quantity comparisons, enabling unambiguous NSE interpretation. Such studies would eliminate the pan evaporation offset issue and provide definitive model performance assessments.

Pan coefficient calibration: empirical determination of region-specific pan coefficients (Kp) relating pan evaporation to ET_0 could establish physically-based conversion relationships, enabling more appropriate performance benchmarks. Seasonal Kp variations for tropical monsoon climates remain inadequately characterized.

Energy balance validation: comparing estimated ET_0 values against independently measured energy balance components (net radiation, soil heat flux, sensible and latent heat) would provide process-based validation independent of pan evaporation proxies.

Multi-site comparative studies: expanding validation to multiple Subak systems with varying elevation, microclimate, and water management practices would assess method transferability and identify location-specific performance variations.

Alternative evaluation metrics: exploring evaluation frameworks specifically designed for comparing related but non-identical quantities, such as modified skill scores or physically-adjusted performance indices, could provide more nuanced performance assessments.

Despite these methodological considerations, the comparative framework employed in this study remains valuable for operational decision-making. The relative performance rankings are robust, as all methods are evaluated against the same observational baseline under identical conditions. The strong positive correlations achieved by FAO Penman-Monteith (r = 0.852) and modified Penman (r = 0.870) demonstrate excellent temporal pattern recognition regardless of absolute magnitude offsets, while Blaney-Criddle's negative correlation (r = -0.782) indicates fundamental pattern failure independent of quantity homogeneity concerns. For practical irrigation scheduling in Subak Balangan and similar systems, the study's operational guidance remains sound: FAO Penman-Monteith provides the most reliable ET_0 estimates when comprehensive data are available, while calibrated modified Penman offers a viable alternative for limited data contexts. The insights into correctable versus structural errors inform method selection and local adaptation strategies regardless of absolute NSE interpretation.

CONCLUSIONS

This comprehensive evaluation of reference evapotranspiration estimation methods in the Subak Balangan irrigation system reveals critical insights for sustainable water resource management in tropical monsoon climates. Multimetric statistical analysis demonstrates that the FAO-56 Penman-Monteith method provides the most accurate performance with excellent pattern recognition (r = 0.852) and lowest magnitude errors (MAE = 0.719 mm/day), validating its international reference standard status for tropical agricultural systems. The systematic underestimation of 0.572 mm/day is physically interpretable through the pan coefficient theory and presents conservative bias preferable for preventing water waste. The Modified Penman method exhibits a critical paradox: highest temporal correlation (r = 0.870), indicating superior seasonal pattern recognition, yet substantial systematic overestimation averaging 27%, producing a negative Nash-Sutcliffe Efficiency. However, diagnostic decomposition reveals this reflects correctable systematic bias rather than structural deficiency, with calibration factor 0.78 transforming performance from unacceptable to good (NSE = 0.684), demonstrating high calibration potential while requiring fewer climate variables than Penman-Monteith. Conversely, the Blaney-Criddle method proves fundamentally unsuitable for tropical monsoon applications despite moderate average errors, exhibiting catastrophic pattern failure manifested by negative correlation (r = -0.782) and misalignment. This structural inadequacy stems from a temperature-only formulation incapable of capturing 44% seasonal evaporation variability dominated by humidity deficit, wind dynamics, and radiation balance in equatorial climates.

The broader implications extend beyond Subak Balangan to similar traditional irrigation systems throughout tropical regions, providing a framework for balancing accuracy with implementation feasibility that is particularly relevant for achieving Sustainable Development Goals 2 and 6 in developing countries where traditional irrigation systems play crucial roles in food security. By implementing these recommendations, irrigation managers can optimize water use efficiency while maintaining crop productivity, potentially reducing dependence on rainfall and mitigating water resource conflicts, thus serving as a model for sustainable water management and climate change adaptation in traditional irrigation systems throughout Indonesia and similar tropical regions. Future research should focus on developing location-specific calibration models, integrating remote sensing data, evaluating economic impacts, and extending comparative analysis to other ET_0 methods to provide more comprehensive methodological guidance for tropical irrigation systems.

Acknowledgments

The authors are grateful to the Community Service Research Institute and Study Center (LPPMPK) Ngurah Rai University. Furthermore, the authors are grateful to the Subak Balangan community, local farmers, and Universitas Ngurah Rai for their support and cooperation, as well as other parties that assisted in completing this study.

REFERENCES

1. Abdelraouf, R.E., El-Shawadfy, M.A., Bakry, A.B., Abdelaal, H.K., El-Shirbeny, M.A., Ragab, R., & Belopukhov, S.L. (2024). Estimating ETO and scheduling crop irrigation using Blaney–Criddle equation when only air-temperature data are available and solving the issue of missing meteorological data in Egypt. *BIO Web of Conferences*, *82*, 1–10. https://doi.org/10.1051/bioconf/20248202020
2. Ahmadi, A., Daccache, A., Sadegh, M., & Snyder, R.L. (2023). Statistical and deep learning models for reference evapotranspiration time series forecasting: A comparison of accuracy, complexity, and data efficiency. *Computers and Electronics in Agriculture*, *215*, 108424. https://doi.org/10.1016/J.COMPAG.2023.108424
3. Allen, R.G., Pereira, L.S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration: Guidelines for romputing rrop water requirements*. Food and Agriculture Organization of the United Nations. https://www.fao.org/4/x0490e/x0490e0b.htm
4. Allen, R.G., Pruitt, W.O., Wright, J.L., Howell, T.A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J.B., Smith, M., Pereira, L.S., Raes, D., Perrier, A., Alves, I., Walter, I., & Elliott, R. (2006). A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. *Agricultural Water Management*, *81*(1–2), 1–22. https://doi.org/10.1016/j.agwat.2005.03.007
5. Alomar, M.K., Hameed, M.M., Al-Ansari, N., & Alsaadi, M.A. (2020). Data-Driven Model for the Prediction of Total Dissolved Gas: Robust Artificial Intelligence Approach. *International Journal of Numerical Methods for Heat \& Fluid Flow*, *30*(6), 2820–2835. https://doi.org/10.1155/2020/6618842
6. Amri, M.F.A., & Heryati, T. (2020). Implementation of sustainable development goals on food security in facing global climate changes. *Universal Journal*, *41*(7–8), 307–307. https://uipmcenter.net/ojs/index.php/journal/article/view/39
7. Anggraheni, E., Zulkarnaain, F., Giardini, P., Maulidina, K., Purbantoro, B., Afifah, R., Muchlis, A., Siswanto, Rustanto, A., Dimyati, M., Zubair, A., Nurlambang, T., Dewanti, R., Ash-Shidiq, I.P., & Susanti, I. (2023). Assessing the reliability of satellite-derived evapotranspiration data using numerical modified penman method at citarum watershed. *Indonesian Journal of Geography*, *55*(2), 213–220. https://doi.org/10.22146/ijg.77725
8. Azizi, M.A., & Sutopo, Y. (2022). Water needs using padi-padi-palawija plants in irrigation area (DI) Sidopangus Regency, Semarang, Indonesia. *Proceedings of the 6th International Conference on Science, Education and Technology (ISET 2020)*, *574*. https://doi.org/10.2991/assehr.k.211125.137
9. Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V.K., & Tuller, M. (2022). Short- and mid-term forecasts of actual evapotranspiration with deep learning. *Journal of Hydrology*, *612*, 128078. https://doi.org/10.1016/J.JHYDROL.2022.128078
10. Chen, H., Zhu, G., Zhang, K., Bi, J., Jia, X., Ding, B., Zhang, Y., Shang, S., Zhao, N., & Qin, W. (2020). Evaluation of evapotranspiration models using different lai and meteorological forcing data from 1982 to 2017. *Remote Sensing*, *12*(15), 2473. https://doi.org/10.3390/rs12152473
11. Chicco, D., Warrens, M.J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, 1–24. https://doi.org/10.7717/PEERJ-CS.623
12. Directorate General of Water Resources. (2015). *Proposed Scheme for Channel Improvement in Pama Palean (Contract No. HK 02 03/BP/PPR/14)*. Ministry of Public Works.
13. Dlouhá, D., Dubovský, V., & Pospíšil, L. (2021). Optimal calibration of evaporation models against penman–monteith equation. *Water (Switzerland)*, *13*(11). https://doi.org/10.3390/w13111484
14. Duc, L., & Sawada, Y. (2023). A signal-processing-based interpretation of the Nash-Sutcliffe efficiency. *Hydrology and Earth System Sciences*, *27*(9), 1827–1839. https://doi.org/10.5194/hess-27-1827-2023
15. Eryani, I.G.A.P., & Jayantari, M.W. (2024). Water conflict analysis in the Balangan Irrigation Area: causes, impacts, and management strategies. *IOP Conf. Series: Earth and Environmental Science 1311*. https://doi.org/10.1088/1755-1315/1311/1/012036
16. Et-Taibi, B., Abid, M.R., Boufounas, E.-M., Morchid, A., Bourhnane, S., & Hame, T.A. (2024). Enhancing water management in smart agriculture: A

cloud and IoT-Based smart irrigation system. *Results in Engineering*, *22*(March), 102283. https://doi.org/10.1016/j.rineng.2024.102283

17. Gandri, L., Fitriani, V., Bowo, C., & Mandala, M. (2024). Comparison of empirical methods to estimated reference evapotranspiration. *Jurnal Ilmiah Rekayasa Pertanian Dan Biosistem*, *12*(2), 177–192. https://doi.org/10.29303/jrpb.v12i2.629
18. Gao, J. (2024). R-Squared (R 2) – How much variation is explained? *Research Methods in Medicine & Health Sciences*, *5*(4), 104–109. https://doi.org/10.1177/26320843231186398
19. Gupta, H.V., Kling, H., Yilmaz, K.K., & Martinez, G.F. (2009). Decomposition of the mean squared error and Nash Sutcliffe efficiency performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003
20. Kim, H., Sim, H., Hong, S., Geem, Z. W., Aksoy, H., Hong, Y., & Yoon, J. (2024). Comparative Evaluation of Evapotranspiration and Optimization Schemes for Green Roof Runoff Simulations Using HYDRUS-1D. *Water* (*Switzerland*), *16*(19). https://doi.org/10.3390/w16192835
21. Kumar, S., Yadav, M., & Kaushik, R. (2025). Ensuring global food security amidst climate change and rapid population growth. *Vigyan Varta*, *6*(1), 1. www.vigyanvarta.com
22. Kusumastuti, D.I., Jokowinarno, D., Putri, R., & Khadijah, L. (2021). The needs and availability analysis of irrigation water of way wayah irigation area. *IOP Conference Series: Earth and Environmental Science*, *739*(1). https://doi.org/10.1088/1755-1315/739/1/012028
23. Liang, Z., Liu, X., Xiong, J., & Xiao, J. (2020). Water allocation and integrative management of precision irrigation: A systematic review. *Water*, *12*(11), 1–23. https://doi.org/10.3390/w12113135
24. Lim, W.H., Roderick, M.L., Hobbins, M.T., Wong, S.C., Groeneveld, P.J., Sun, F., & Farquhar, G.D. (2012). The aerodynamics of pan evaporation. *Agricultural and Forest Meteorology*, *152*(C), 31–43. https://doi.org/10.1016/j.agrformet.2011.08.006
25. Liu, X., Xu, C., Zhong, X., Li, Y., Yuan, X., & Cao, J. (2017). Comparison of 16 models for reference crop evapotranspiration against weighing lysimeter measurement. *Agricultural Water Management*, *184*, 145–155. https://doi.org/10.1016/j.agwat.2017.01.017
26. Melsen, L.A., Puy, A., Torfs, P.J.J.F., & Saltelli, A. (2025). The rise of the Nash-Sutcliffe efficiency in hydrology. *Hydrological Sciences Journal*, *70*(8), 1248–1259. https://doi.org/10.1080/02626667.2025.2475105
27. Mendoza, C.J., & Quiñones, A.J.P. (2021). *Revista Brasileira de Engenharia Agrícola e Ambiental Reference evapotranspiration estimation by different methods for the sucroenergy sector of Colombia 1 Estimativa da evapotranspiração de referência por diferentes métodos para o setor sucroenergético d*. 583–590. https://www.researchgate.net/publication/351328321_Reference_evapotranspiration_estimation_by_different_methods_for_the_sucroenergy_sector_of_Colombia
28. Nusantara, D.A.D., & Nadiar, F. (2020). Using ANN to Evaluate the Climate Data that High Affect on Calculate Daily Potential Evapotranspiration with Modified-Penman Method in the Tropical Regions. *Journal of Physics: Conference Series*, *1569*(4). https://doi.org/10.1088/1742-6596/1569/4/042028
29. Obilor, E.I., & Amadi, E.C. (2018). Test for Significance of Pearson's Correlation Coefficient (r). *International Journal of Innovative Mathematics, Statistics & Energy Policies*, *6*(1), 11–23. https://www.researchgate.net/publication/323522779_Test_for_Significance_of_Pearson's_Correlation_Coefficient
30. Pandey, V., Srivastava, P.K., Das, P., & Behera, M.D. (2021). Irrigation water demand estimation in Bundelkhand region using the variable infiltration capacity model. *Agricultural Water Management: Theories and Practices*, 331–347. https://doi.org/10.1016/B978-0-12-812362-1.00016-3
31. Parwata. (2021). *After nearly 20 years of drought, Subak Balangan has experienced drought, leading Badung officials to visit the location*. 8 Mei. https://www.balipost.com/news/2021/05/08/190897/Hampir-20-Tahun-Subak-Balangan...html
32. Saleem, A., Anwar, S., Nawaz, T., Fahad, S., Saud, S., Rahman, T.U., Khan, M.N.R., & Nawaz, T. (2024). Securing a sustainable future: the climate change threat to agriculture, food security, and sustainable development goals. *Journal of Umm Al-Qura University for Applied Sciences*, *0123456789*. https://doi.org/10.1007/s43994-024-00177-3
33. Sambo, U., & Sule, B. (2023). Impact of Climate Change on Food Security in Northern Nigeria. *Green and Low-Carbon Economy*, *2*(1), 49–61. https://doi.org/10.47852/bonviewglce3202560
34. Sasireka, K., Reddy, C.J.M., Reddy, C.C., & Ramakrishnan, K. (2017). Evaluation and recalibration of empirical constant for estimation of reference crop evapotranspiration against the Modified Penman method. *IOP Conference Series: Earth and Environmental Science*, *80*(1), 12062. https://doi.org/10.1088/1755-1315/80/1/012062
35. Saura, R.B.D., & Andante, R.J.M. (2018). Detection of cyanide in freshwater fishes relative to sex dimorphism using landmark-based geometric morphometrics in Agusan del Sur, Philippines. *International Journal of Biosciences (IJB)*, *12*(2), 177–193. https://doi.org/10.12692/ijb/12.2.177-193
36. Schober, P., Boer, C., & Schwarte, L.A. (2018).

Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia \& Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

37. Sharafi, S., Ghaleni, M.M., & Scholz, M. (2023). Comparison of predictions of daily evapotranspiration based on climate variables using different data mining and empirical methods in various climates of Iran. *Heliyon*, *9*(2), e13245. https://doi.org/10.1016/j.heliyon.2023.e13245
38. Suryadi, E., Ruswandi, D., Dwiratna, S., & Prawiranegara, B.M.P. (2019). Crop Water Requirements Analysis Using Cropwat 8.0 Software in Maize Intercropping with Rice and Soybean. *International Journal on Advanced Science, Engineering and Information Technology*, *9*(4), 1364–1370. https://doi.org/10.18517/ijaseit.9.4.6868
39. Toušková, J., Falatkova, K., & Šipek, V. (2025). Estimating Potential and Reference Evapotranspiration in the Central European Region: The Challenge of Model Selection. *Water Resources Management*, *39*(11), 5911–5927. https://doi.org/10.1007/s11269-025-04233-3
40. Wadman, W. (2023). Wadman W. Efficient irrigation and water management: Ensuring sustainable agriculture and conservation of water resources Efficient irrigation and water management: Ensuring sustainable agriculture and conservation of water resources. *J Agric Sci Bot*, *7*(3), 177. https://doi.org/10.35841/2591-7366-7.3.177