

Article

Using Artificial Intelligence Algorithms to Estimate and Short-Term Forecast the Daily Reference Evapotranspiration with Limited Meteorological Variables

Shih-Lun Fang ^{1,†} , Yi-Shan Lin ^{1,†}, Sheng-Chih Chang ², Yi-Lung Chang ², Bing-Yun Tsai ² and Bo-Jein Kuo ^{1,3,*} 

¹ Department of Agronomy, National Chung Hsing University, Taichung 40227, Taiwan; alanfang19941214@gmail.com (S.-L.F.); lin90036@gmail.com (Y.-S.L.)

² Taiwan Seed Improvement and Propagation Station, Taichung 426017, Taiwan; james7309@tss.gov.tw (S.-C.C.); niceyilung@tss.gov.tw (Y.-L.C.); bingyun@tss.gov.tw (B.-Y.T.)

³ Smart Sustainable New Agriculture Research Center (SMARTer), Taichung 40227, Taiwan

* Correspondence: bjkuo@dragon.nchu.edu.tw; Tel.: +886-4-22840777 (ext. 201)

† These authors contributed equally to this work.

Abstract: The reference evapotranspiration (ET_0) information is crucial for irrigation planning and water resource management. While the Penman-Monteith (PM) equation is widely recognized for ET_0 calculation, its reliance on numerous meteorological parameters constrains its practical application. This study used 28 years of meteorological data from 18 stations in four geographic regions of Taiwan to evaluate the effectiveness of an artificial intelligence (AI) model for estimating PM-calculated ET_0 using limited meteorological variables as input and compared it with traditional methods. The AI models were also employed for short-term ET_0 forecasting with limited meteorological variables. The findings suggested that AI models performed better than their counterpart methods for ET_0 estimation. The artificial neural network using temperature, solar radiation, and relative humidity as input variables performed best, with the correlation coefficient (r) ranging from 0.992 to 0.998, mean absolute error (MAE) ranging from 0.07 to 0.16 mm/day, and root mean square error (RMSE) ranging from 0.12 to 0.25 mm/day. For short-term ET_0 forecasting, the long short-term memory model using temperature, solar radiation, and relative humidity as input variables was the best structure to forecast four-day-ahead ET_0 , with the r ranging from 0.608 to 0.756, MAE ranging from 1.05 to 1.28 mm/day, and RMSE ranging from 1.35 to 1.62 mm/day. The percentage error of this structure was within $\pm 5\%$ for most meteorological stations over the one-year test period, underscoring the potential of the proposed models to deliver daily ET_0 forecasts with acceptable accuracy. Finally, the proposed estimating and forecasting models were developed in regional and variable-limited scenarios, making them highly advantageous for practical applications.

Keywords: artificial neural network; long short-term memory; reference evapotranspiration; Penman-Monteith equation; limited meteorological variables



Citation: Fang, S.-L.; Lin, Y.-S.; Chang, S.-C.; Chang, Y.-L.; Tsai, B.-Y.; Kuo, B.-J. Using Artificial Intelligence Algorithms to Estimate and Short-Term Forecast the Daily Reference Evapotranspiration with Limited Meteorological Variables. *Agriculture* **2024**, *14*, 510. <https://doi.org/10.3390/agriculture14040510>

Academic Editors: Miltiadis Iatrou, Christos Karydas and Panagiotis Tziachris

Received: 15 February 2024

Revised: 20 March 2024

Accepted: 20 March 2024

Published: 22 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the escalation of the global population and the concomitant rise in water demand, effectively managing limited water resources while sustaining productivity has become an important issue in agriculture today [1]. Consequently, assessing crop water demand has become crucial to achieving a precise allocation of water resources [1,2]. Crop water demand can be obtained from the product of reference evapotranspiration (ET_0) and crop-specific coefficients [2,3]. Moreover, ET_0 can serve to characterize local climatic conditions or aid in calculating drought indices for drought monitoring [4–6]. Therefore, an accurate estimation of ET_0 is essential for irrigation planning and water resource management.

ET_0 is defined as the evapotranspiration of a hypothetical crop with a plant height of 0.12 m, a fixed surface resistance of 70 s/m, and an albedo of 0.23 [7]. Although

ET_0 can be measured by a lysimeter, the use of this instrument in practical applications is typically constrained by cost and operational intricacies, making it more suitable for research purposes [8,9]. Conversely, the approach of using the meteorological data recorded by weather stations to establish ET_0 equations through mathematical and physical theories is a more practical avenue for application [10].

Among the existing ET_0 equations, the Penman-Monteith (PM) equation stands out as the most renowned [11,12]. The PM equation is deduced based on energy balance and water vapor diffusion, and it also considers crop canopy resistance and aerodynamic resistance. Since the laws of physics do not change from location to location, the physics-based PM equation is sufficient as a basis for estimating ET_0 globally when high-quality meteorological data can be obtained [13]. Therefore, the PM equation was recommended as a standard method for estimating ET_0 in the FAO-56 report of the Food and Agriculture Organization (FAO) of the United Nations in 1998. Subsequent sensitivity analyses and regional assessments have confirmed the applicability of the PM equation across diverse environments [10,13]. However, the application of the PM equation requires many meteorological parameters including, solar radiation (R_s), mean air temperature (T), mean relative humidity (RH), and mean wind speed at 2 m above ground (u_2). The construction and maintenance of corresponding sensors can be expensive. In addition, it is difficult to collect reliable meteorological data in some regions, or some parameters cannot be achieved [14]. Therefore, developing methods that could accurately estimate ET_0 using a subset of meteorological variables had value from either a cost or practical perspective.

Three strategies have been developed to address the absence of requisite parameters for the PM equation, including using PM-alternative equations, estimating the lacking meteorological variable(s), and estimating the ET_0 based on artificial intelligence (AI) algorithms. The FAO recommends initially estimating the lacking meteorological variable and then plugging the estimate into the PM equation [13], an approach known as the reduced-set PM (RPM) [15,16]. There are also PM-alternative equations such as Hargreaves-Samani (HS), Makkink, and Turc equations that require fewer meteorological variables and are easier to compute than the PM equation. However, the performance of these equations varies from location to location, so they generally require local calibration, which limits the applicability of alternative models [17]. Recently, various AI-based substitution approaches to the PM equation have been reported.

Since evapotranspiration is a nonlinear and complex phenomenon, deriving a formula capable of encapsulating all relevant physical phenomena becomes a challenge [18]. Therefore, neural networks have emerged as a prominent tool, owing to their capacity to delineate input-output relationships without necessitating an intricate understanding of underlying physical mechanisms [19–21]. Antonopoulos and Antonopoulos [22] employed an artificial neural network (ANN) with a limited set of input variables to estimate daily ET_0 . Their ANN models yielded values of correlation coefficient (r) ranging from 0.952 to 0.978 and 0.910 to 0.956 when utilizing three input variables (T -RH- R_s) and two input variables (T - R_s or T -RH), respectively. Ferreira et al. [10] evaluated the efficacy of ANN and a support vector machine in daily ET_0 estimation in Brazil using temperature and RH or temperature alone, concluding that adding temperature and RH data from four previous days as input for the ANN model had the best performance. Chen et al. [23] evaluated the performance of three deep learning methods, i.e., deep neural network, temporal convolution neural network (TCN), and long short-term memory (LSTM), for daily ET_0 estimation with incomplete meteorological variables, showing that the TCN and LSTM models consistently outperformed PM-alternative equations across all tested scenarios. With recent advancements in AI, a variety of hybrid algorithms have been proposed to increase the accuracy of stand-alone algorithms [17]. Sharma et al. [24] proposed a convolutional neural network-LSTM (CNN-LSTM) for ET_0 estimation using limited meteorological variables, surpassing the performance of existing PM-alternative equations such as HS and Makkink. Xing et al. [25] introduced a hybrid model named D-LSTM, which integrated the deep belief network (DBN) module for feature extraction from meteorological data and the LSTM

module for sequential feature processing. The D-LSTM model synergistically leveraged the strengths of DBN and LSTM, achieving superior performance in daily ET_0 estimation with incomplete meteorological variables on the Loess Plateau.

In addition to estimating ET_0 from past and current periods, forecasting ET_0 in advance can help optimize irrigation scheduling [26,27], for example, to improve timely irrigation scheduling and water resource management [28]. Although monthly or yearly ET_0 forecasting aids in the middle- to long-term irrigation planning, daily ET_0 forecasting is particularly advantageous for short-term irrigation management, especially for systems requiring high-frequency irrigation [27]. However, daily ET_0 forecasting is more complex than monthly or yearly forecasting since the daily ET_0 data fluctuates more [17]. Many studies have employed forecasted meteorological data as inputs into the ET_0 equation (e.g., HS) to achieve ET_0 forecasting [15,27]. However, the efficacy of such a method hinges on the forecasting accuracy of the meteorological data [11,29], as any biases in meteorological variable forecasting can result in systematic errors in daily ET_0 forecasts [15,29]. Another way to address this challenging task is to use AI algorithms, e.g., LSTM, to deal with the sequential data. Ferreira and da Cunha [26] utilized CNN-LSTM with past ET_0 values and meteorological variables as inputs to forecast ET_0 seven days in advance. Roy et al. [30] incorporated past ET_0 values as inputs into the bi-directional LSTM to forecast five-day-ahead ET_0 . However, the use of past ET_0 values may not be suitable for scenarios with limited variables, as it necessitates ET_0 calculation beforehand. Yin et al. [31] developed the hybrid bi-directional LSTM (Bi-LSTM) with three meteorological inputs (maximum temperature, minimum temperature, and sunshine duration), which outperformed the adjusted HS method for short-term (1–7-day lead time) daily ET_0 forecasting. Zhang et al. [16] demonstrated the effectiveness of the LSTM model in the short-term (1–7 days ahead) daily ET_0 forecasting across diverse climate zones in China. Despite the growing interest in forecasting ET_0 , the literature on this topic with data over 20 years is limited [16,26,30,31].

Taiwan has abundant rainfall, around 2000 mm per year, but the amount of available water per person is only one-sixth of the global average. This discrepancy arises from the geographical characteristics of Taiwan, characterized by short and steep rivers, leading to rapid rainfall discharge into the sea with uneven rainfall patterns across the regions and seasons. This variability makes the utilization and management of water resources more difficult. Considering that agricultural irrigation accounts for more than 60% of Taiwan's water resources usage, accurate estimation and forecasting of ET_0 could help in the determination of the water requirements for crop cultivation to make informed decisions regarding water management. Therefore, this study used about three decades of meteorological data collected from stations across Taiwan to (1) evaluate the effectiveness of ANN in estimating ET_0 using limited meteorological variables as inputs compared to the two traditional methods; (2) use LSTM and CNN-LSTM to conduct short-term ET_0 forecasting with limited variables based solely on historical meteorological data.

2. Materials and Methods

2.1. Study Area and Meteorological Data Collection

Taiwan Island is situated in the northwestern Pacific Ocean, spanning latitudes between 22–25° N and longitudes between 120–122° E. Taiwan Island can be broadly divided into four geographic regions, namely, the northern, central, southern, and eastern regions. Northern and central Taiwan have a subtropical monsoon climate, whereas southern Taiwan has a tropical monsoon climate. Due to the numerous high mountains in Taiwan, some areas have the characteristics of a temperate climate.

Daily meteorological data, including T, maximum air temperature, minimum air temperature, RH, R_s , u_2 , and mean atmospheric pressure were collected from 18 meteorological stations across the four geographic regions between 1 January 1995 and 31 December 2022. These meteorological stations are located within the primary crop cultivation zones of each geographic region (Figure 1). The original meteorological data were downloaded from

the website of the Central Weather Administration of Taiwan (<https://agr.cwa.gov.tw/>), accessed on 18 January 2024).

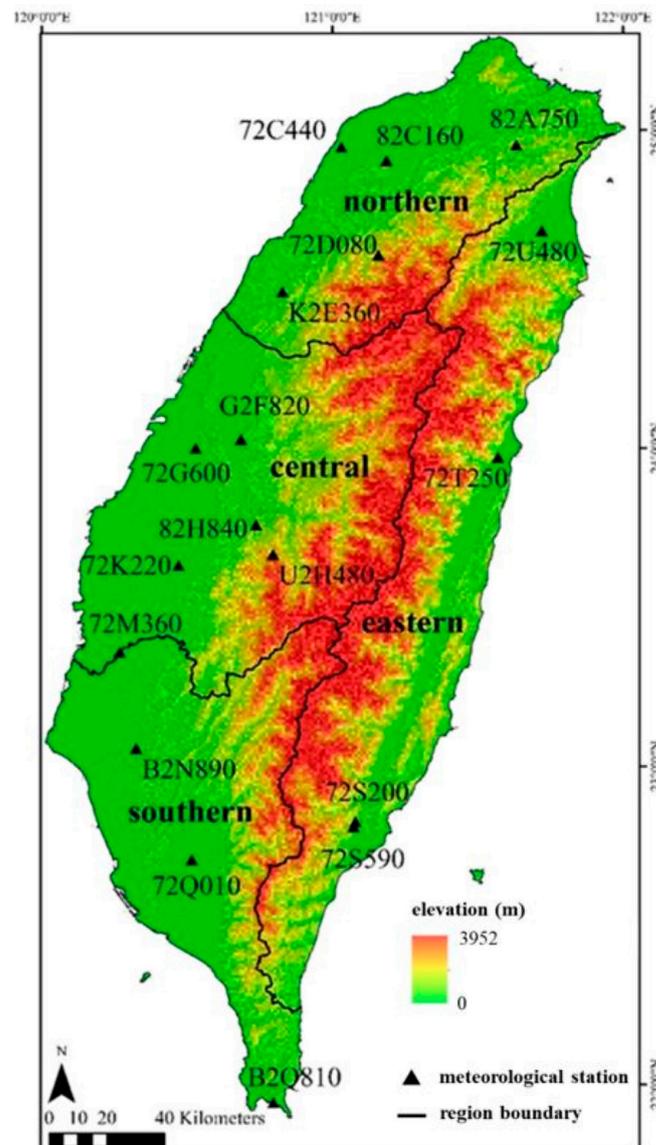


Figure 1. Spatial distribution of the 18 meteorological stations utilized in this study.

2.2. Data Processing and PM-ET₀ Calculation

The total missing rate of the original dataset was about 10%. The missing values in the dataset were imputed with the same day-of-the-year average [32] for each station. After missing value imputation, the PM equation (Equation (1)) [7] was used to calculate daily ET₀. A summary of the meteorological characteristics of the studied stations is presented in Table 1.

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

$$e_s = 0.6108 \times \exp\left(\frac{17.27T}{T + 273.3}\right) \quad (2)$$

$$e_a = e_s \times \frac{RH}{100} \quad (3)$$

$$\Delta = 4098 \times \frac{e_s}{(T + 273.3)^2} \tag{4}$$

$$\gamma = 0.00163 \times \frac{P}{2.45} \tag{5}$$

where R_s represents the solar radiation ($\text{MJ}/\text{m}^2/\text{day}$); G denotes the soil heat flux density ($\text{MJ}/\text{m}^2/\text{day}$); T is the mean air temperature ($^{\circ}\text{C}$); u_2 indicates the mean wind speed at 2 m height (m/s); e_s stands for the saturation vapor pressure (kPa); e_a is the actual vapor pressure (kPa); Δ denotes the slope of the vapor pressure curve ($\text{kPa}/^{\circ}\text{C}$); γ indicates the psychrometric constant ($\text{kPa}/^{\circ}\text{C}$); RH represents the mean relative humidity (%); and P is the mean atmospheric pressure (kPa) [33,34]. Because G is typically negligible compared to R_s , it is assumed to be zero on a daily scale.

Table 1. Summary of the meteorological characteristics of each station. The mean and standard deviation (SD) of each station were calculated using the daily data from 1995 to 2022.

Region	Station	T ($^{\circ}\text{C}$)		RH (%)		u_2 (m/s)		R_s ($\text{MJ}/\text{m}^2/\text{day}$)		ET_0 (mm/day)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Northern	72C440	22.27	5.36	80.85	9.40	3.94	1.95	12.97	7.39	4.19	2.23
	72D080	17.32	4.80	88.31	9.64	0.50	0.24	9.83	6.03	2.74	1.76
	82A750	20.06	5.33	85.94	10.01	1.45	1.21	10.43	8.00	3.11	2.38
	82C160	21.56	5.53	82.30	9.86	2.89	1.50	11.76	7.15	3.71	2.23
	K2E360	22.23	5.39	82.30	9.19	2.36	1.32	13.18	6.34	4.04	2.02
Central	72G600	23.39	4.93	80.54	7.83	1.99	0.92	11.60	5.32	3.72	1.64
	72K220	23.38	4.81	81.22	7.40	1.87	0.82	12.35	5.00	3.88	1.57
	72M360	23.59	4.83	82.76	7.89	3.02	1.22	15.03	5.86	4.51	1.77
	82H840	21.52	4.38	87.25	11.13	0.91	0.46	12.69	5.26	3.64	1.59
	G2F820	23.52	4.73	81.67	7.31	2.21	0.92	12.53	5.72	3.94	1.78
U2H480	17.05	3.55	89.49	6.32	1.06	0.35	9.24	3.97	2.49	1.08	
Southern	72Q010	24.98	3.80	78.70	8.45	1.45	0.63	12.35	6.40	4.02	1.95
	B2N890	23.71	4.49	82.31	7.08	1.82	0.82	12.19	5.18	3.80	1.58
	B2Q810	25.34	3.11	77.25	7.93	4.12	1.65	16.40	6.54	5.47	1.80
Eastern	72S200	22.50	4.02	81.18	8.02	1.51	0.47	8.72	4.85	2.88	1.54
	72S590	22.43	3.76	86.26	8.85	1.30	0.69	7.84	5.31	2.48	1.58
	72T250	22.92	4.24	81.50	6.77	1.13	0.54	10.56	6.67	3.34	2.01
	72U480	22.45	4.99	83.44	8.36	1.82	0.76	11.64	7.67	3.63	2.36

T: mean air temperature; RH: mean relative humidity; u_2 : mean wind speed at 2 m height; R_s : solar radiation; ET_0 : reference evapotranspiration.

2.3. Daily ET_0 Estimation under Variable-Limited Conditions

2.3.1. Use of PM-Alternative Equations

When it is difficult to obtain certain meteorological variables, other equations that do not require as many meteorological variables as the PM equation can be used to estimate ET_0 [33,35]. The HS (Equation (6)) [36], Makkink (Equation (7)) [37], and Turc (Equation (8)) [35] are three commonly used PM-alternative equations. The HS equation only requires observed temperature data, while the Makkink equation requires both T and R_s data. Conversely, the Turc equation needs T , R_s , and RH data.

$$\text{ET}_0 = 0.0023 \frac{R_a}{2.45} (T + 17.8) \sqrt{(T_{\text{max}} - T_{\text{min}})} \tag{6}$$

$$\text{ET}_0 = 0.61 \left(\frac{R_s}{2.45} \frac{\Delta}{\Delta + \gamma} \right) - 0.12 \tag{7}$$

$$ET_0 = \begin{cases} 0.013(23.88R_s + 50) \left(\frac{T}{T+15} \right), & \text{if RH} \geq 50\% \\ 0.013(23.88R_s + 50) \left(\frac{T}{T+15} \right) \left(1 + \frac{50-RH}{70} \right), & \text{if RH} < 50\% \end{cases} \quad (8)$$

where T is the mean air temperature (°C); T_{min} and T_{max} are the minimum and maximum temperature (°C), respectively; R_s denotes solar radiation (MJ/m²/d); Δ represents the slope of the vapor pressure curve (kPa/°C) and can be achieved by Equation (4); γ indicates the psychrometric constant (kPa/°C) and can be calculated by Equation (5); RH stands for the mean relative humidity (%); and R_a is extraterrestrial radiation (MJ/m²/day) which can be obtained by latitude and day of year (DoY) as follows [34]:

$$R_a = \frac{1440}{\pi} G_{sc} d_r^2 [\omega_s \sin(\text{lat}) \sin(\delta) + \cos(\text{lat}) \cos(\delta) \sin(\omega_s)] \quad (9)$$

where G_{sc} is the solar constant and equals 0.0820 MJ/m²/min; d_r stands for the Earth-Sun inverse relative distance; d_r² = 1 + 0.033cos($\frac{2\pi}{365}$ DoY); lat represents latitude (rad) (negative for the southern hemisphere); δ denotes the solar declination (rad), δ = 0.409sin($\frac{2\pi}{365}$ DoY - 1.39); and ω_s is the sunset hour angle (rad), ω_s = arcos[-tan(lat)tan(δ)].

2.3.2. Estimation of Lacking Meteorological Variables in RPM Approach

Meteorological variables that are readily accessible and reliably measured can be used to estimate the lacking meteorological variables required for the PM equation. When the observed RH data are unattainable, the e_a can be estimated by Equation (10). Equation (11) offers an estimation method for R_s in the absence of observed radiation data. The regional or global average wind speed (2 m/s) can be used as a substitute when the u₂ is not measured [33,38].

$$E_a = 0.611 \times \exp\left(\frac{17.27T_{\min}}{T_{\min} + 237.3}\right) \quad (10)$$

$$R_s = k_{R_s} \sqrt{R_a(T_{\max} - T_{\min})} \quad (11)$$

where e_a is the actual vapor pressure (kPa); T_{min} and T_{max} are the minimum and maximum temperature (°C), respectively; k_{R_s} denotes the adjustment coefficient that needs to be locally calibrated; and R_a represents extraterrestrial radiation (MJ/m²/day) calculated by Equation (9).

This study used 28-year meteorological data to calibrate the average wind speed (used to substitute the u₂) and k_{R_s} (required to estimate R_s), which are specific for each geographic region (Table 2). The different values in each geographic region reflect the distinction in climate conditions among the four regions.

Table 2. Results of the calibration of the average wind speed and k_{R_s}.

Region	\hat{k}_{R_s}	\hat{u}_2
Northern	0.13	2.23
Central	0.12	1.84
Southern	0.15	2.46
Eastern	0.10	1.44

\hat{k}_{R_s} : estimated adjustment coefficient k_{R_s}; \hat{u}_2 : estimated mean wind speed at 2 m height.

2.3.3. ANN Models

The typical ANN architecture comprises an input layer for data introduction, one or more hidden layers for data processing, and an output layer for generating the final results [17]. The basic operation of ANN is as follows: initially, data are inputted into the network’s input layer, where each input variable is represented by a neuron. Subsequently, the information in each neuron is weighted and summed, with the resulting weighted sum serving as input for the subsequent layer. Finally, the output is converted through

an activation function such as the logistic and hyperbolic tangent commonly used in hydrological research [19]. After the model is constructed, the ANN measures the difference between the model output and the reference value through a loss function, iteratively adjusting the weights until it finds a set of weights that minimize the loss function value.

This study employed feed-forward multilayer perceptron ANNs and an Adam training algorithm. The ratio of training data commonly used in machine learning is 60% to 80% [10,16,31]; therefore, the 28 years of daily data from each meteorological station were randomly split into three parts for model training (70%), validation (15%), and testing (15%), respectively. The training set was used to develop the ANN model and optimize the model parameters. The validation set served to assess overfitting during the training stage, and the testing set evaluated the generalizability of the developed model. Next, data from all stations within the same geographic region were pooled to construct ANN models for each region.

The following hyperparameter combinations were evaluated: the number of epochs was 50, and the batch size was 32. The hyperbolic tangent function was used as an activation function [19], and the type of loss function was the mean square error. The model becomes more complex as the number of hidden layers and units increases, thereby increasing the computational cost and causing overfitting. Additionally, increasing the number of neurons is not always good [39]; therefore, this study assessed the number of hidden layers ranging from 1 to 3 layers and the number of hidden units ranging from 10 to 200 with a step size of 10. The learning rate was set to 0.001. Additionally, a patience value of 5 was set [40,41], indicating that if there is no improvement in the validation error after 5 epochs, the training process would be stopped to prevent overfitting. The T, RH, and R_s were used as input variables, and the PM-calculated ET₀ was the output variable (Figure 2). The input variables were standardized using Equation (12) to ensure a uniform contribution from each feature before training and testing [31,42]. Four combinations of input variables were explored: (1) T, (2) T and R_s, (3) T and RH, and (4) T, R_s, and RH. The ANNs were implemented using Tensorflow2.6 based on Python (version 3.7), and the standardization was implemented using the StandardScaler function in the sklearn.preprocessing package based on the scikit-learn module. The best combination of the number of hidden layers and hidden units was determined by grid search. The final hyperparameter configurations of the ANN models using different input variable combinations for each geographic region are summarized in Table 3. The data are standardized as follows:

$$x_s = \frac{x_o - \bar{x}}{SD} \tag{12}$$

where x_o and x_s are values before and after the standardization, respectively. \bar{x} and SD are the mean and standard deviation of the data, respectively.

Table 3. The hyperparameter values with different input combinations for each geographic region.

Inputs Variables	Number of Hidden Layers/Number of Hidden Units			
	Northern	Central	Southern	Eastern
T	3/170	2/200	2/140	2/110
T, R _s	1/190	1/170	1/180	1/180
T, RH	1/140	1/170	1/160	1/120
T, R _s , RH	1/180	1/200	1/150	1/160

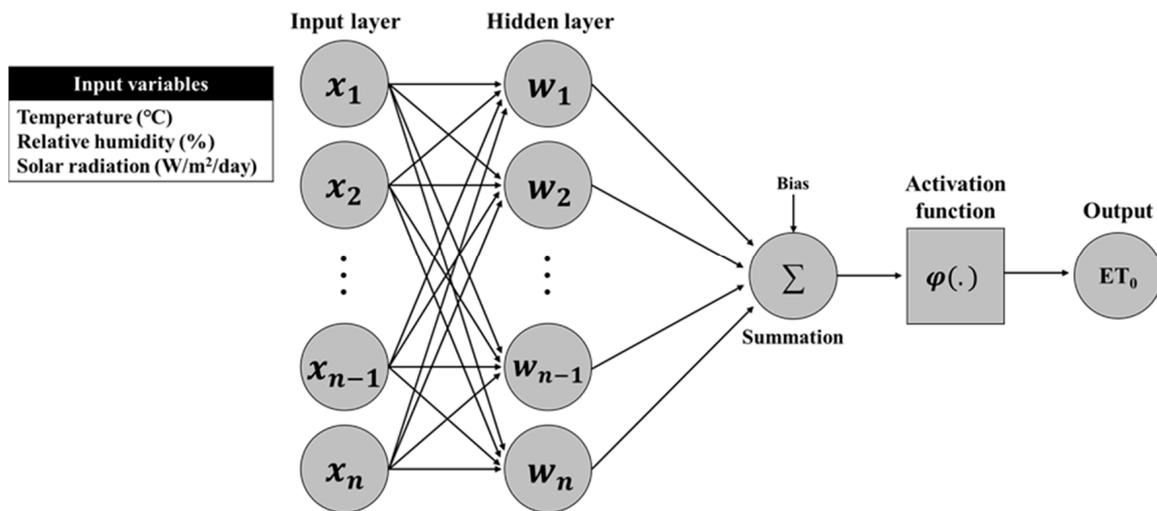


Figure 2. The general architecture of the artificial neural network used in the study. ET_0 represents the reference evapotranspiration. This Figure was adapted from Goyal et al. [17].

2.4. Short-Term Forecasting Daily ET_0 under Variable-Limited Conditions

Since LSTM can capture and model temporal dependencies, it is widely used for processing time-series data [43]. A neuron in the LSTM model comprises a cell state and three gates: the input gate (i_t), forget gate (f_t), and output gate (O_t), enabling selective retention or modification of information. The general LSTM structure is shown in Figure 3a and its computing steps can be described as the following equations:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (13)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (14)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (15)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (16)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (17)$$

$$h_t = O_t \odot \tanh(C_t) \quad (18)$$

where \tilde{C}_t is the cell state candidate; C_t denotes the new cell state; h_t is the hidden state vector; x_t represents the input vector; $W_i, W_f, W_C,$ and W_o stand for the weight matrices; $b_i, b_f, b_C,$ and b_o are the bias vectors; $\sigma(\cdot)$ is the sigmoid function; $\tanh(\cdot)$ represents the hyperbolic tangent function; and \odot denotes the element-wise product [31,40,44].

In CNN-LSTM architecture, the convolution layer is employed for extracting features from input data, and the extracted features are subsequently fed into LSTM layers for processing information with sequential features. The convolution layers extract salient features automatically, with a one-dimensional (1D) convolutional layer typically used for sequential data or time series [26]. In the CNN-LSTM model, an input layer is followed by a 1D convolutional layer; subsequently, the output of the convolutional layer undergoes further processing through LSTM layers, and finally, the model concludes with an output layer in the form of a dense layer.

A 90%–5%–5% split [40] was used to allow the models to learn complex patterns, so the 28 years of daily data from each meteorological station were sequentially divided into three parts: training (1 January 1995 to 13 March 2020), validation (14 March 2020 to 6 August 2021), and testing (7 August 2021 to 31 December 2022). Subsequently, the

data of all stations belonging to the same geographic region were pooled to construct ET_0 forecasting models for each region. In this study, the past meteorological variables were used as inputs to forecast four-day-ahead ET_0 , and three combinations of input variables were examined: (1) T and R_s , (2) T and RH, and (3) T, R_s , and RH. A standardization process (Equation (12)) was used to ensure an equal contribution from each feature. The model structure comprised one 1D convolutional layer, one 1D max pooling layer, one to three LSTM layers, and one densely connected NN layer (Figure 3b). The following hyperparameter combinations were evaluated: the number of epochs was 50, and the batch size was 32. The hyperbolic tangent function was used as the activation function [19], and the mean square error was used as the loss function. The number of hidden units was set from 10 to 200 with a step size of 10, the learning rate was set to 0.001, and the optimizer used the Adam algorithm. This study chose a 90%–5%–5% split percentage to establish models, but this extreme split could increase the risk of overfitting. Therefore, a dropout rate of 0.4 [40] and a patience value of 5 for early stopping [40,41] were set to prevent overfitting. The models were developed using Tensorflow2.6 based on Python (version 3.7), and the standardization was implemented using the StandardScaler function. In this study, the best combination of the number of hidden layers and hidden units was determined by grid search. The final hyperparameter configurations for the LSTM and CNN-LSTM models using different combinations of input variables for each geographic region are summarized in Table 4.

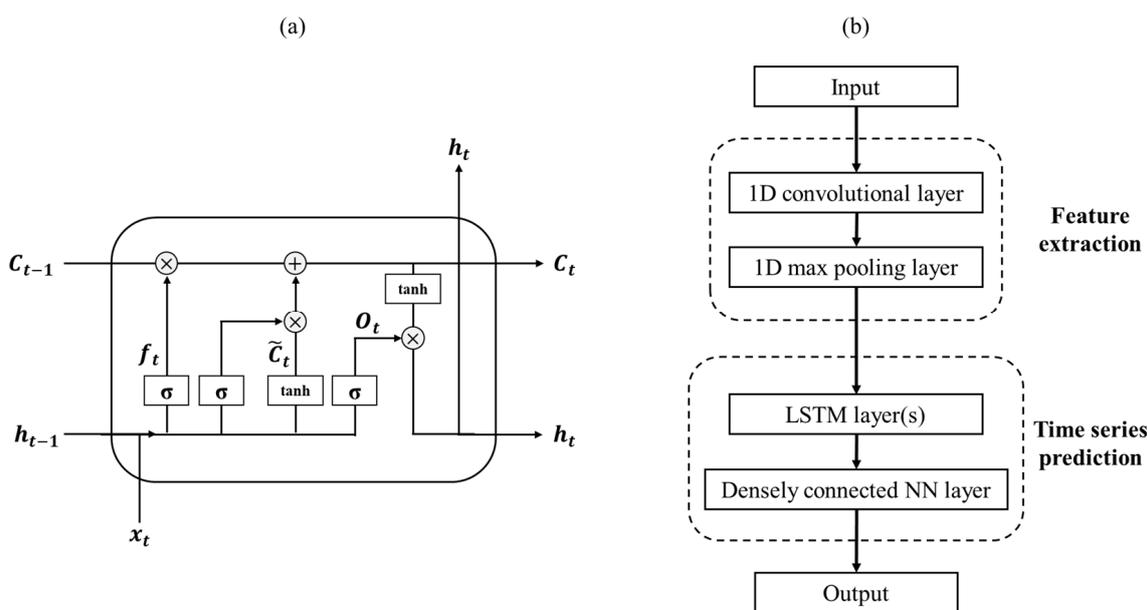


Figure 3. (a) General structure of an LSTM cell. (b) The CNN-LSTM architecture used in the study. (a) was adapted from Lin et al. [40].

Table 4. The hyperparameter values of the ET_0 forecasting models with different input combinations for each geographic region.

Algorithm	Inputs Variables	Number of LSTM Layers/Number of LSTM Units			
		Northern	Central	Southern	Eastern
LSTM	T, R_s	1/160	3/110	3/150	1/180
	T, RH	3/70	2/160	3/200	2/70
	T, R_s , RH	1 /110	2/110	3/200	1/190
CNN-LSTM	T, R_s	1/170	3/180	3/180	3/200
	T, RH	2/10	3/190	3/180	1/30
	T, R_s , RH	1/190	2/180	3/140	2/120

2.5. Performance Comparison Criteria

The performance of each method was evaluated using three standard statistical metrics: the r , mean absolute error (MAE), and root mean square error (RMSE). The value of r ranges between -1 and 1 , serving as an indicator to evaluate the linear relationship between the predicted and observed values. MAE and RMSE quantify the bias between the predicted value and the observed value, with values ranging from 0 to infinity. A higher r , along with a lower MAE and RMSE, signifies a more accurate prediction.

$$r = \frac{\sum_{i=1}^n (o_i - \bar{o})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^n (o_i - \bar{o})^2} \sqrt{\sum_{i=1}^n (p_i - \bar{p})^2}} \tag{19}$$

$$MAE = \frac{\sum_{i=1}^n |o_i - p_i|}{n} \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (o_i - p_i)^2}{n}} \tag{21}$$

where n is the number of observations in the testing set for the AI algorithms (ANN, LSTM, and CNN-LSTM) and the total number of 28-year data for other methods; o_i is the observed value (the PM-calculated ET_0); p_i is the predicted (estimated or forecasted) ET_0 value by the other methods; and \bar{o} and \bar{p} are the mean of o_i and p_i , respectively.

3. Results and Discussion

3.1. Evaluation of ET_0 Estimation Approaches under Variable-Limited Conditions

3.1.1. Performance of ET_0 Estimation Methods

This study used 28 years of meteorological data from 18 stations distributed across the four geographic regions of Taiwan to evaluate the efficacy of various ET_0 estimation approaches under variable-limited conditions (Table 5). Regarding the alternative equations, the r between HS-estimated ET_0 and PM-calculated ET_0 ranged from 0.444 to 0.735 , with MAE ranging from 0.95 to 1.41 mm/day and RMSE from 1.22 to 1.85 mm/day. For the Makkink equation, the r ranged from 0.972 to 0.990 , MAE ranged from 1.33 to 1.93 mm/day, and RMSE ranged from 1.53 to 2.08 mm/day. The Turc equation performed the best among the three PM-alternative equations, with the r ranging from 0.984 to 0.997 , MAE ranging from 0.50 to 0.84 mm/day, and RMSE ranging from 0.68 to 1.03 mm/day.

Table 5. Assessment of daily ET_0 estimation methods across different geographic regions. The results highlighted in bold indicate a superior performance.

Method's Name	Variables Need Observed	Northern			Central			Southern			Eastern		
		r	MAE	RMSE									
HS	T, T_{max}, T_{min}	0.735	1.22	1.56	0.699	0.95	1.22	0.444	1.41	1.85	0.685	1.22	1.50
Makkink	T, R_s	0.978	1.50	1.75	0.980	1.52	1.67	0.972	1.93	2.08	0.990	1.33	1.53
Turc	T, R_s, RH	0.989	0.68	0.89	0.992	0.59	0.74	0.984	0.84	1.03	0.997	0.50	0.68
RPM (T)	T	0.735	2.42	3.04	0.698	2.44	2.84	0.357	2.87	3.39	0.719	2.12	2.72
RPM (T, u_2)	T, u_2	0.674	2.47	3.07	0.631	2.46	2.85	0.468	2.99	3.47	0.645	2.13	2.72
RPM (T, R_s)	T, R_s	0.975	0.33	0.49	0.980	0.28	0.36	0.953	0.41	0.59	0.990	0.21	0.27
RPM (T, RH)	T, RH	0.804	2.44	3.03	0.704	2.59	2.96	0.610	2.81	3.28	0.706	2.17	2.77
RPM (T, u_2, R_s)	T, u_2, R_s	0.976	0.28	0.48	0.983	0.24	0.33	0.952	0.38	0.61	0.992	0.18	0.25
RPM (T, u_2, RH)	T, u_2, RH	0.655	2.40	3.00	0.616	2.56	2.94	0.527	2.82	3.28	0.614	2.15	2.76
RPM (T, R_s, RH)	T, R_s, RH	0.995	0.14	0.24	0.996	0.09	0.16	0.990	0.18	0.27	0.998	0.07	0.12
ANN (T)	T	0.701	1.25	1.56	0.616	1.04	1.32	0.543	1.36	1.68	0.731	1.04	1.32
ANN (T, R_s)	T, R_s	0.979	0.27	0.45	0.983	0.21	0.31	0.977	0.30	0.43	0.991	0.19	0.25
ANN (T, RH)	T, RH	0.828	0.94	1.23	0.744	0.87	1.12	0.710	1.12	1.40	0.803	0.88	1.15
ANN (T, R_s, RH)	T, R_s, RH	0.994	0.14	0.24	0.996	0.10	0.15	0.992	0.16	0.25	0.998	0.07	0.12

Regarding the RPM approach, it does not perform well when only considering the temperature, i.e., RPM (T), as evidenced by r ranging from 0.357 to 0.735, MAE ranging from 2.12 to 2.87 mm/day, and RMSE ranging from 2.72 to 3.39 mm/day (Table 5). For the combination of two variables, RPM (T, R_s) performed the best, with r ranging from 0.953 to 0.990, MAE ranging from 0.21 to 0.41 mm/day, and RMSE ranging from 0.27 to 0.59 mm/day; in contrast, RPM (T, u_2) performed the worst, with r ranging from 0.468 to 0.674, MAE ranging from 2.13 to 2.99 mm/day, and RMSE ranging from 2.72 to 3.47 mm/day. In the RPM model with three variables, the accuracy of estimating ET_0 is the worst when the R_s variable is not included in the model, with r of RPM (T, u_2 , RH) ranging from 0.527 to 0.655, MAE ranging from 2.15 to 2.82 mm/day, and RMSE ranging from 2.76 to 3.28 mm/day. However, the lack of u_2 has the most negligible impact on accuracy within the three-variable model, as evidenced by r of RPM (T, R_s , RH) ranging from 0.990 to 0.998, MAE ranging from 0.07 to 0.18 mm/day, and RMSE ranging from 0.12 to 0.27 mm/day.

In terms of the application of ANN to establish the ET_0 estimation models, the combinations of (T, R_s) and (T, R_s , RH) demonstrated a favorable performance, with r ranging from 0.977 to 0.998, MAE ranging from 0.07 to 0.30 mm/day, and RMSE ranging from 0.12 to 0.45 mm/day (Table 5). The performance of the proposed ANN models is similar to that of ANN using the same input variables reported in Antonopoulos and Antonopoulos [22]. In addition, the performance of the ANN (T) model is comparable to that of the HS equation, with r ranging from 0.543 to 0.731, MAE ranging from 1.04 to 1.36 mm/day, and RMSE ranging from 1.32 to 1.68 mm/day.

3.1.2. Comparison of Different ET_0 Estimation Methods and Inputs Combination

The results in Table 5 revealed that irrespective of the approach employed, using the combination of T, R_s , and RH (e.g., Turc, RPM (T, R_s , RH) and ANN (T, R_s , RH)) achieved the best results, characterized by r above 0.984 with MAE and RMSE values below 0.84 and 1.03 mm/day, respectively. Notably, both RPM (T, R_s , RH) and ANN (T, R_s , RH) slightly outperformed the Turc equation. Conversely, the performance of each approach is poor when only considering the temperature in the model (e.g., HS, RPM (T), and ANN (T)), especially in southern Taiwan. For the case when only temperature data were available, the HS equation and ANN (T) exhibited equally poor performance, but with noticeably lower MAE values compared to RPM (T). Yang et al. [15] conducted a comparison between the HS equation and RPM (T), concluding that the HS equation outperformed RPM in subtropical regions. Similarly, Córdova et al. [45] reported that the HS equation performed slightly better than RPM (T), but still yielded an unsatisfactory result.

The results in Table 5 indicate that accurate ET_0 estimates can still be achieved even without u_2 . Therefore, as mentioned by Córdova et al. [45], the absence of u_2 was not a major source of error in humid climates. The role of u_2 in ET_0 calculation is subject to two main perspectives: some argue that it is a decisive factor due to the potential for measurement errors, while others contend that u_2 does not have much impact on ET_0 [17]. Nevertheless, Fisher et al. [46] stated that wind speed plays an important role in modeling the complex nonlinear behavior of ET_0 . Furthermore, Makwana et al. [47] found that the performance of the ANN model could be significantly improved in windy regions by including wind speed and T_{max} , with r and MAE values of 0.91 and 0.59 mm/day, respectively.

Temperature and radiation are considered essential inputs to estimate ET_0 [22,24]; therefore, including observed R_s in the ET_0 calculation generally yields satisfactory results, except that the Makkink equation exhibits higher MAE and RMSE values (Table 5). RH is an important variable in estimating ET_0 in humid regions [17], such as Taiwan. Consequently, the Makkink equation, which does not incorporate RH information, performs poorly. However, RPM (T, R_s) also lacks observed RH data, but the information about RH is estimated by Equation (10), hence it can still perform well. Conversely, methods lacking R_s observations perform poorly (Table 5). Even though the RPM approach estimates the absent R_s information through Equation (11), its performance remains unsatisfactory. Córdova et al. [45] observed that estimating R_s based on Equation (11) yielded poor results in humid condi-

tions for the RPM approach. Sentelhas et al. [12] discovered that when the actual R_s falls below $20 \text{ MJ/m}^2/\text{day}$, Equation (11) tends to systematically overestimate radiation. Given that Taiwan exhibits a humid climate with R_s values less than $20 \text{ MJ/m}^2/\text{day}$ (Table 1), this could account for the suboptimal performance of RPM in this study.

This study aimed to evaluate the effectiveness of ANN in estimating ET_0 using limited meteorological variables. Traore et al. [48] noted that the ANN model fed with only temperature data performed better than the HS equation, which included some radiation information by using R_a as inputs. In this study, the ANN (T) model yielded r ranging from 0.543 to 0.731, MAE ranging from 1.04 to 1.36 mm/day, and RMSE ranging from 1.32 to 1.68 mm/day (Table 5). Notably, the ANN (T) model exhibited superior performance compared to RPM (T) and demonstrated comparable efficacy to the HS equation. In scenarios involving two input variables, the ANN (T, RH) model exhibited r ranging from 0.710 to 0.828, MAE ranging from 0.87 to 1.12 mm/day, and RMSE ranging from 1.12 to 1.40 mm/day; the ANN (T, RH) performed better than RPM (T, RH). On the other hand, the ANN (T, R_s) model demonstrated r ranging from 0.977 to 0.991, MAE ranging from 0.19 to 0.30 mm/day, and RMSE ranging from 0.25 to 0.45 mm/day, surpassing both RPM (T, R_s) and the Makkink equation. When T, RH, and R_s data are available, the ANN (T, R_s , RH) model exhibited r ranging from 0.992 to 0.998, MAE ranging from 0.07 to 0.16 mm/day, and RMSE ranging from 0.12 to 0.25 mm/day. The ANN (T, R_s , RH) model performed better than the Turc equation and showed comparable performance to RPM (T, R_s , RH). These findings underscore the superior performance of ANN models compared to traditional methods utilizing the same inputs. Additionally, the performance of ANN using T and R_s as input variables is comparable with the models using more input variables, as mentioned by Jain et al. [49].

3.2. Performance of AI Algorithms for the Short-Term Forecasting ET_0 with Limited Variables

This study used 28 years of meteorological data from 18 stations in four geographic regions of Taiwan to construct LSTM and CNN-LSTM models for short-term ET_0 forecasting with limited meteorological variables. When considering scenarios with two input variables, the LSTM (T, RH) model exhibited an r ranging from 0.159 to 0.711, MAE ranging from 1.29 to 1.60 mm/day, and RMSE ranging from 1.60 to 2.01 mm/day (Table 6). Notably, the LSTM (T, RH) performed better than CNN-LSTM (T, RH) in the majority of regions. Similarly, the LSTM (T, R_s) model yielded the values of r ranging from 0.609 to 0.755, MAE ranging from 1.04 to 1.31 mm/day, and RMSE ranging from 1.35 to 1.62 mm/day, outperforming the CNN-LSTM (T, R_s) model. Considering T, RH, and R_s data simultaneously, the LSTM (T, R_s , RH) model attained the values of r ranging from 0.608 to 0.756, MAE ranging from 1.05 to 1.28 mm/day, and RMSE ranging from 1.35 to 1.62 mm/day, exhibiting a superior performance compared to the CNN-LSTM (T, R_s , RH) model (Table 6). Moreover, the performance of the proposed models was consistent between the training (Table S1) and testing phases (Table 6) indicating that overfitting did not occur.

Table 6. Testing performance of short-term ET_0 forecasting models in each geographic region. Bold characters indicate better results among the models.

Model's Name	Inputs Variables	Northern			Central			Southern			Eastern		
		r	MAE	RMSE									
LSTM (T, R_s)	T, R_s	0.746	1.31	1.62	0.743	1.04	1.35	0.609	1.23	1.62	0.755	1.13	1.47
LSTM (T, RH)	T, RH	0.711	1.50	1.82	0.606	1.32	1.60	0.159	1.60	2.01	0.691	1.29	1.62
LSTM (T, R_s , RH)	T, R_s , RH	0.751	1.28	1.61	0.744	1.05	1.35	0.608	1.23	1.62	0.756	1.12	1.47
CNN-LSTM (T, R_s)	T, R_s	0.642	1.55	1.87	0.617	1.29	1.59	0.341	1.45	1.84	0.683	1.31	1.64
CNN-LSTM (T, RH)	T, RH	0.612	1.63	1.93	0.476	1.46	1.77	0.336	1.74	2.15	0.557	1.51	1.87
CNN-LSTM (T, R_s , RH)	T, R_s , RH	0.742	1.32	1.63	0.733	1.08	1.37	0.595	1.24	1.64	0.742	1.17	1.51

For CNN-LSTM models, the expectation is that CNN layers can extract features from input data, while LSTM layers can capture time patterns to enhance the model's capability and achieve higher performance. However, in the present study, the CNN-LSTM models performed worse than the LSTM models. This finding contrasts with those of Barzegar et al. [50] and Yin et al. [31], who reported that integrated or hybridized algorithms outperformed standalone algorithms for ET_0 forecasting. Moreover, while hybrid algorithms can enhance the model performance, the extent of improvement may not always be substantial [51]. Additionally, hybrid algorithms typically necessitate tuning more hyperparameters, making them time-consuming during model deployment [17]. Zhang et al. [16] compared LSTM to three novel AI models and found that the LSTM model marginally outperformed in short-term daily ET_0 forecasting across diverse climate conditions. Considering the computational efficiency and model performance, the LSTM (T, R_s , RH) was regarded as the best structure in this study.

In evaluating the LSTM (T, R_s , RH) model across various geographic regions for forecasting four-day-ahead ET_0 at each meteorological station, the overall trend of the daily ET_0 forecasting was consistent with the daily ET_0 values calculated by the PM equation (Figures 4–7), underscoring the forecasting ability of proposed models. During the test period lasting more than one year, the disparity between the cumulative ET_0 forecasted by the LSTM (T, R_s , RH) model and the observed ET_0 ranged from -166.05 to 68.12 mm, with most differences within $\pm 5\%$, except for stations 72D080 and 72S590 (Table 7). The poor performance of the forecasting models at these two stations may be due to their differing meteorological characteristics from most stations in the same region (Table 1). Interestingly, there was a large difference between predicted and observed values in autumn (September–November) (Figures 4–7), which is different from the previous studies that reported that winter and summer were more prone to forecasting errors [52,53]. This might be due to Taiwan's climate conditions suffering the large meteorological fluctuations affected by typhoons, southwest monsoons, and northeast monsoons [54]. The model performance could be increased by using the day of the year as an additional input variable [26] to provide information about seasons or grouping the data into different seasons to establish a specific model for each season [55]. Future studies should explore these strategies to refine ET_0 forecasting models.

Table 7. Forecasting result of LSTM (T, R_s , RH) models and the cumulative values of model-forecasted ET_0 and the observed PM-calculated ET_0 of each station for the test period (7 August 2021 to 31 December 2022).

Region	Station	Observed ET_0 (mm)	Forecasted ET_0 (mm)	Error ^a / Percentage Error ^b
Northern	72C440	2435.03	2333.41	$-101.61 / -4.17\%$
	72D080	1694.40	1594.58	$-99.82 / -5.89\%$
	82A750	1679.82	1747.94	$68.12 / 4.06\%$
	82C160	2206.86	2200.66	$-6.19 / -0.28\%$
	K2E360	2194.46	2163.84	$-30.62 / -1.40\%$
Central	72G600	2626.29	2525.82	$-100.47 / -3.83\%$
	72K220	2516.21	2433.42	$-82.79 / -3.29\%$
	72M360	2689.00	2567.05	$-121.95 / -4.54\%$
	82H840	2131.62	2075.59	$-56.04 / -2.63\%$
	G2F820	2614.64	2525.49	$-89.15 / -3.41\%$
	U2H480	1396.68	1407.67	$10.99 / 0.79\%$
Southern	72Q010	3221.07	3063.86	$-157.21 / -4.88\%$
	B2N890	2493.22	2554.38	$61.16 / 2.45\%$
	B2Q810	2897.23	2845.22	$-52.01 / -1.80\%$
Eastern	72S200	2051.92	1974.23	$-77.70 / -3.79\%$
	72S590	2000.07	1834.02	$-166.05 / -8.30\%$
	72T250	1963.31	1956.38	$-6.93 / -0.35\%$
	72U480	1914.17	1933.81	$19.64 / 1.03\%$

^a Error = Forecasted value – Observed value. ^b Percentage error = $\frac{\text{Forecasted value} - \text{Observed value}}{\text{Observed value}} \times 100\%$.

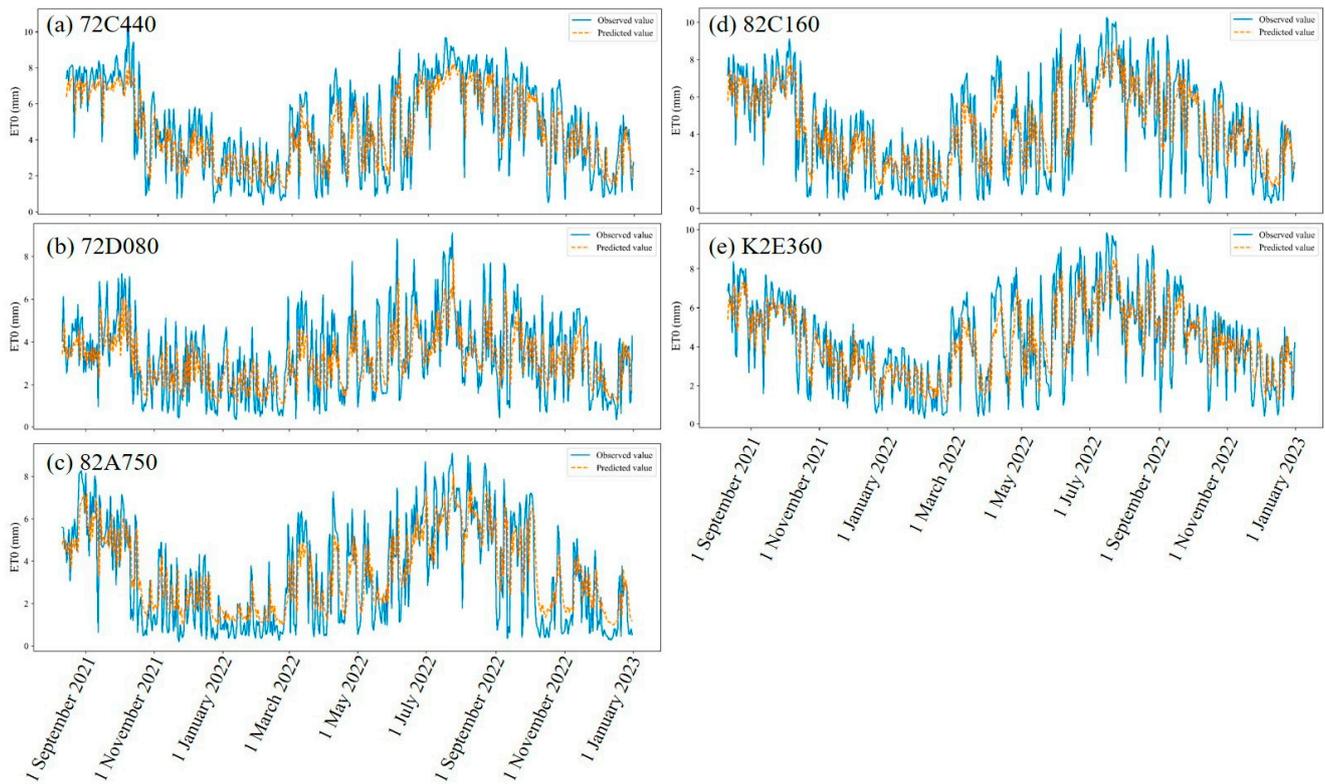


Figure 4. Comparison between the observed values of PM-calculated ET_0 and the four-day-ahead forecasted values of the LSTM (T, R_s, RH) model during the test period at the following stations in northern Taiwan: (a) 72C440, (b) 72D080, (c) 82A750, (d) 82C160, and (e) K2E360.

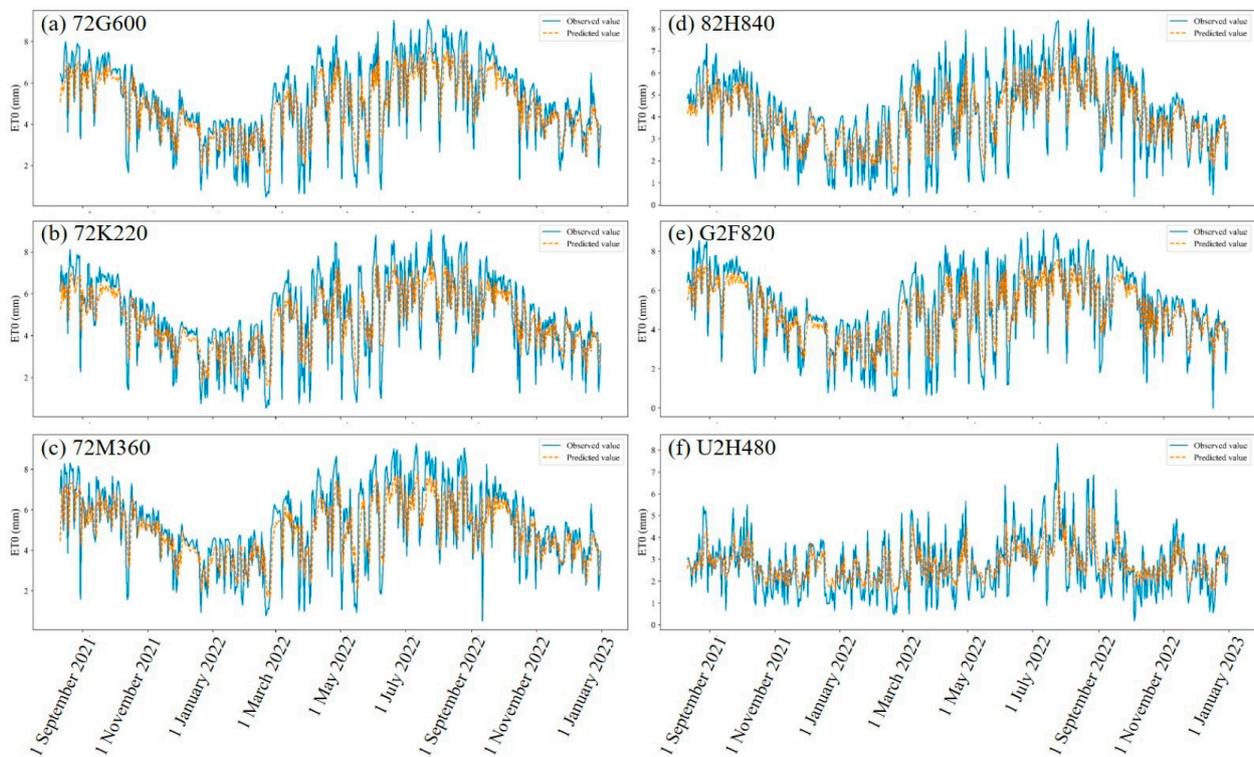


Figure 5. Comparison between the observed values of PM-calculated ET_0 and the four-day-ahead forecasted values of the LSTM (T, R_s, RH) model during the test period at the following stations in central Taiwan: (a) 72G600, (b) 72K220, (c) 72M360, (d) 82H840, (e) G2F820, and (f) U2H480.

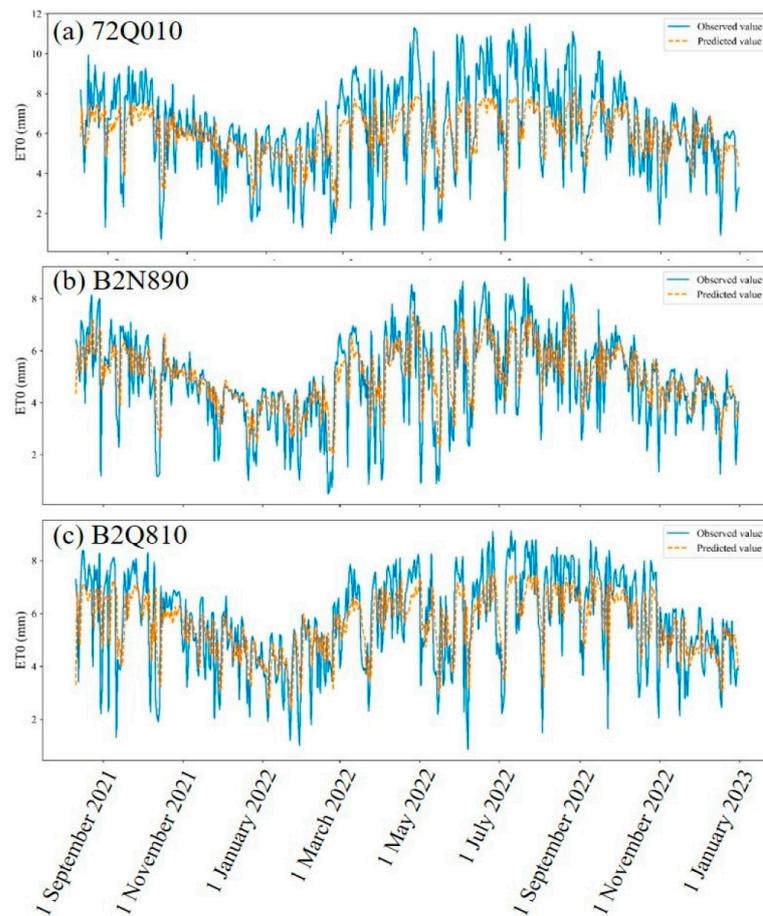


Figure 6. Comparison between the observed values of PM-calculated ET_0 and the four-day-ahead forecasted values of the LSTM (T , R_s , RH) model during the test period at the following stations in southern Taiwan: (a) 72Q010, (b) B2N890, and (c) B2Q810.

3.3. Advantages and Limitations of the Proposed Models

This study shows that the AI algorithm can sufficiently estimate and even forecast ET_0 using only a few meteorological variables but its forecasting capabilities could be improved and tested over longer periods to increase the reliability. Many studies caution that AI models remain empirical, implying that a model developed for one location may not generalize well to the other locations [16,19]. Increasing the data variability for modeling may increase the generalizability of AI models [10,19]. This study developed ET_0 estimating and forecasting models in regional scenarios, i.e., trained with pooled data from several meteorological stations, to achieve a higher generalization capacity. These regional models are more readily applicable in practical settings, as a single model can suffice for an entire region. However, previous studies pointed out that the regional models exhibit lower accuracy than those developed in localized scenarios due to dissimilarities in the training data [10,19,56]. To consider the generalizability and performance, the stations could be grouped based on meteorological characteristics rather than geographical locations, and then a specific model was built for each cluster [10,57]. This strategy is expected to improve the performance of proposed regional models at specific stations.

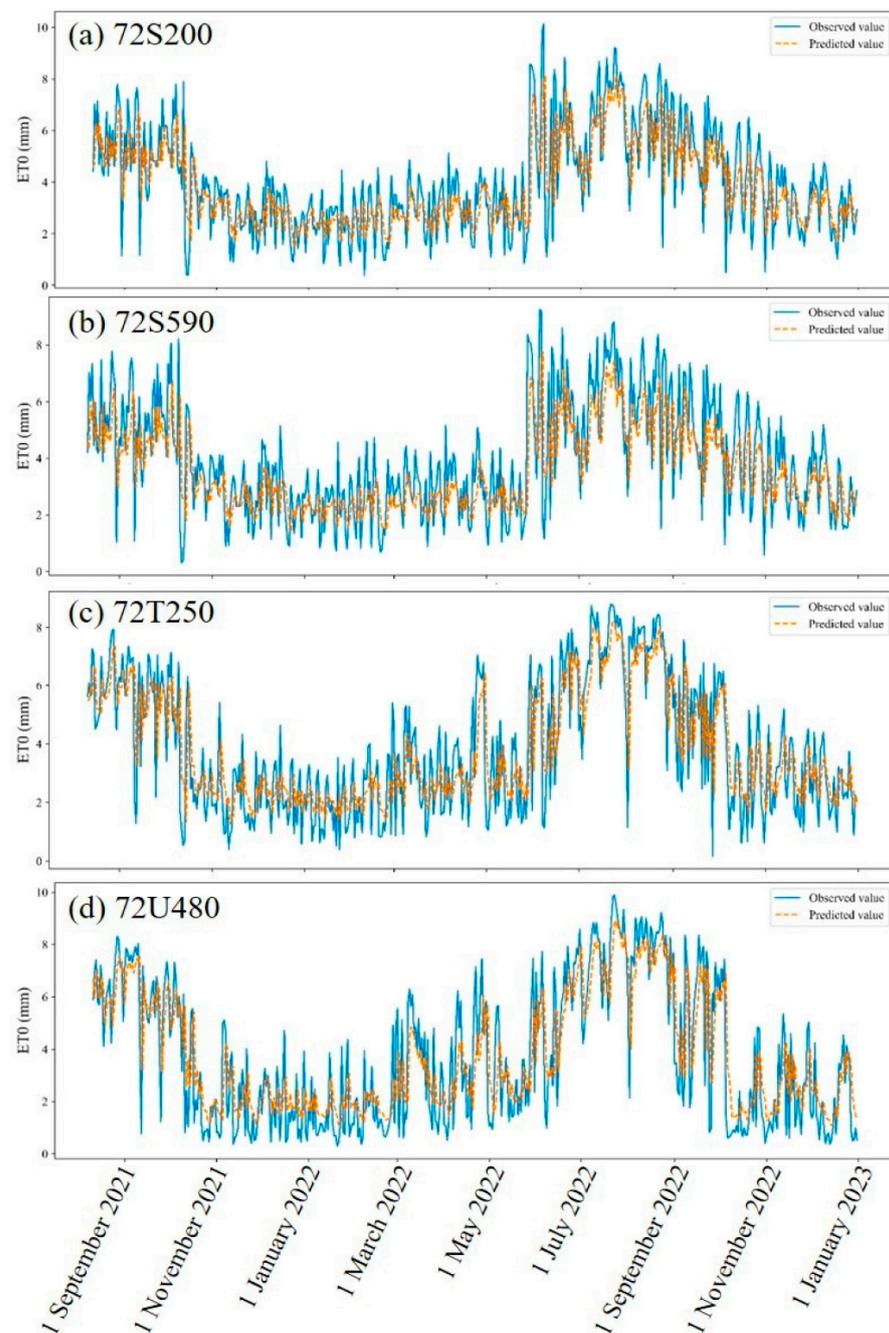


Figure 7. Comparison between the observed values of PM-calculated ET_0 and the four-day-ahead forecasted values of the LSTM (T , R_s , RH) model during the test period at the following stations in eastern Taiwan: (a) 72S200, (b) 72S590, (c) 72T250, and (d) 72U480.

It must be mentioned that the ANN and LSTM endeavor to mimic the underlying evapotranspiration process through the link weights and thresholds. However, the black-box nature of these models impedes their ability to elucidate the relationship between input and output through explicit expressions [17,19]. Consequently, they cannot supplant physical mechanisms [16,19]. In addition, climate change affects evapotranspiration, not only in Taiwan [58,59] but also in other regions [60–67] worldwide. Consequently, a “fixed” forecasting model may not make accurate predictions in the future [17], so the forecasting models should be trained with long-term data and continuously adjusted by incorporating new data through dynamic modeling [17,68]. Furthermore, greenhouse cultivation is rapidly growing and as the water demand cannot be met through rainfall, it must be

supplied through irrigation. However, irrigation in greenhouses is often based on a fixed schedule or the grower's experience, often leading to insufficient irrigation or excessive waste of water resources [69]. Utilizing evapotranspiration information to determine the appropriate timing and amount of irrigation can greatly enhance the precision of greenhouse cultivation. Most studies estimated or forecasted evapotranspiration under open-field conditions [70], but applying these models to greenhouse environments resulted in discrepancies due to the vastly different conditions [70,71]; therefore, future studies should focus more on estimating and forecasting evapotranspiration within greenhouses.

4. Conclusions

Since ET_0 is related to the quantification of crop water demand, the information on ET_0 is useful for irrigation planning and water resources management. However, the standard method for calculating ET_0 , the PM equation, is primarily limited by the costs associated with collecting numerous meteorological parameters. Considering that AI algorithms are known for modeling complex relationships without requiring an in-depth understanding of underlying physical processes, they have great potential to obtain ET_0 information with fewer variables. The results of this study revealed that ANN models outperformed the traditional methods (RPM and PM-alternative equations) with the same inputs for ET_0 estimation. Particularly, the ANN (T, R_s , RH) exhibited the highest accuracy (r was 0.992 to 0.998, MAE was 0.07 to 0.16 mm/day, and RMSE was 0.12 to 0.25 mm/day), while the ANN utilizing T and R_s yielded satisfactory results (r was 0.977 to 0.991, MAE was 0.19 to 0.30 mm/day, and RMSE was 0.25 to 0.45 mm/day). For the short-term ET_0 forecasting, LSTM models performed better than CNN-LSTM models, with LSTM (T, R_s , RH) demonstrating the best performance in forecasting four-day-ahead ET_0 (r ranging from 0.608 to 0.756, MAE ranging from 1.05 to 1.28 mm/day, and RMSE ranging from 1.35 to 1.62 mm/day). Over the one-year test period, the error percentage of LSTM (T, R_s , RH) models was within $\pm 5\%$ for most meteorological stations but these models can be tested over longer periods to increase their reliability. Moreover, the forecasting capabilities can be further improved by using the day of the year as an additional input variable or grouping/clustering data. As the proposed ANN and LSTM models were developed in regional and variable-limited scenarios, they offer practical utility across the entire region without necessitating the inclusion of all meteorological variables. In addition, these models may be applied to other regions that have similar conditions to Taiwan. Future studies can focus more on estimating and forecasting evapotranspiration within greenhouses. Additionally, it is important to consider the impact of climate change. Accurate ET_0 estimation and forecasting models enable the agricultural sector to determine irrigation water requirements promptly or in advance. As the models are refined and their applicability is expanded, they could significantly benefit precise water managing and irrigation scheduling in crop production.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agriculture14040510/s1>, Table S1. Training performance of short-term ET_0 forecasting models in each geographic region. Bold characters indicate better results among the models.

Author Contributions: Conceptualization, S.-L.F.; Data curation, S.-L.F., Y.-S.L., S.-C.C., Y.-L.C. and B.-Y.T.; Funding acquisition, B.-J.K.; Investigation, S.-L.F., S.-C.C., Y.-L.C. and B.-Y.T.; Methodology, S.-L.F. and Y.-S.L.; Writing—original draft, S.-L.F. and Y.-S.L.; Writing—review & editing, B.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by NSTC 112-2634-F-005-002—project Smart Sustainable New Agriculture Research Center (SMARTer).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
CNN-LSTM	Convolution neural network long- short-term memory
e_a	Actual vapor pressure
ET_0	Reference evapotranspiration
HS equation	Hargreaves-Samani equation
LSTM	Long short-term memory
MAE	Mean absolute error
PM equation	Penman-Monteith equation
r	Correlation coefficient
R_a	Extraterrestrial radiation
R_s	Solar radiation
RMSE	Root mean square error
RH	Relative humidity
RPM	Reduced-set Penman-Monteith
T	Mean air temperature
T_{max}	Maximum temperature
u_2	Mean wind speed at 2 m above ground

References

- Sharma, V.; Irmak, S. Mapping spatially interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirements in Nebraska: Part I. Precipitation and reference evapotranspiration. *Trans. ASABE* **2012**, *55*, 907–921. [\[CrossRef\]](#)
- Sharma, V.; Irmak, S. Mapping spatially interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirements in Nebraska: Part II. Actual crop evapotranspiration and net irrigation requirements. *Trans. ASABE* **2012**, *55*, 923–936. [\[CrossRef\]](#)
- Rahman, M.A.; Ali, M.; Mojid, M.A.; Anjum, N.; Haq, M.E.; Kainose, A.; Dissanayaka, K.D.C.R. Crop coefficient, reference crop evapotranspiration and water demand of dry-season Boro rice as affected by climate variability: A case study from northeast Bangladesh. *Irrig. Drain.* **2023**, *72*, 148–165. [\[CrossRef\]](#)
- Dai, A. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. *J. Geophys. Res.* **2011**, *116*, D12115. [\[CrossRef\]](#)
- Paulo, A.A.; Rosa, R.D.; Pereira, L.S. Climate trends and behavior of drought indices based on precipitation and evapotranspiration in Portugal. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 1481–1491. [\[CrossRef\]](#)
- McEvoy, D.J.; Huntington, J.L.; Abatzoglou, J.T.; Edwards, L.M. An evaluation of multi-scalar drought indices in Nevada and Eastern California. *Earth Interact.* **2012**, *16*, 1–18. [\[CrossRef\]](#)
- Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; FAO: Rome, Italy, 1998; Volume 56.
- Allen, R.G.; Pereira, L.S.; Howell, T.A.; Jensen, M.E. Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agric. Water Manag.* **2011**, *98*, 899–920. [\[CrossRef\]](#)
- Landeras, G.; Ortiz-Barredo, A.; López, J.J. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agric. Water Manag.* **2008**, *95*, 553–565. [\[CrossRef\]](#)
- Ferreira, L.B.; da Cunha, F.F.; de Oliveira, R.A.; Fernandes Filho, E.I. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM—A new approach. *J. Hydrol.* **2019**, *572*, 556–570. [\[CrossRef\]](#)
- Cai, J.B.; Liu, Y.; Lei, T.W.; Pereira, L.S. Estimating reference evapotranspiration with the FAO Penman-Monteith equation using daily weather forecast messages. *Agric. For. Meteorol.* **2007**, *145*, 22–35. [\[CrossRef\]](#)
- Sentelhas, P.C.; Gillespie, T.J.; Santos, E.A. Evaluation of FAO Penman-Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. *Agric. Water Manag.* **2010**, *97*, 635–644. [\[CrossRef\]](#)
- Pereira, L.S.; Allen, R.G.; Smith, M.; Raes, D. Crop evapotranspiration estimation with FAO56: Past and future. *Agric. Water Manag.* **2015**, *147*, 4–20. [\[CrossRef\]](#)
- Tabari, H.; Talaee, P.H. Local calibration of the Hargreaves and Priestley-Taylor equations for estimating reference evapotranspiration in arid and cold climates of Iran based on the Penman-Monteith model. *J. Hydrol. Eng.* **2011**, *16*, 837–845. [\[CrossRef\]](#)
- Yang, Y.; Cui, Y.; Bai, K.; Luo, T.; Dai, J.; Wang, W.; Luo, Y. Short-term forecasting of daily reference evapotranspiration using the reduced-set Penman-Monteith model and public weather forecasts. *Agric. Water Manag.* **2019**, *211*, 70–80. [\[CrossRef\]](#)
- Zhang, L.; Zhao, X.; Zhu, G.; He, J.; Chen, J.; Chen, Z.; Traore, S.; Liu, J.; Singh, V.P. Short-term daily reference evapotranspiration forecasting using temperature-based deep learning models in different climate zones in China. *Agric. Water Manag.* **2023**, *289*, 108498. [\[CrossRef\]](#)

17. Goyal, P.; Kumar, S.; Sharda, R. A review of the Artificial Intelligence (AI) based techniques for estimating reference evapotranspiration: Current trends and future perspectives. *Comput. Electron. Agric.* **2023**, *209*, 107836. [[CrossRef](#)]
18. Wen, X.; Si, J.; He, Z.; Wu, J.; Shao, H.; Yu, H. Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resour. Manag.* **2015**, *29*, 3195–3209. [[CrossRef](#)]
19. Kumar, M.; Raghuvanshi, N.S.; Singh, R. Artificial neural networks approach in evapotranspiration modeling: A review. *Irrig. Sci.* **2011**, *29*, 11–25. [[CrossRef](#)]
20. Adeloye, A.J.; Rustum, R.; Kariyama, I.D. Neural computing modeling of the reference crop evapotranspiration. *Environ. Model. Softw.* **2012**, *29*, 61–73. [[CrossRef](#)]
21. Tikhamarine, Y.; Malik, A.; Souag-Gamane, D.; Kisi, O. Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration. *Environ. Sci. Pollut. Res.* **2020**, *27*, 30001–30019. [[CrossRef](#)]
22. Antonopoulos, V.Z.; Antonopoulos, A.V. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Comput. Electron. Agric.* **2017**, *132*, 86–96. [[CrossRef](#)]
23. Chen, Z.; Zhu, Z.; Jiang, H.; Sun, S. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *J. Hydrol.* **2020**, *591*, 125286. [[CrossRef](#)]
24. Sharma, G.; Singh, A.; Jain, S. A hybrid deep neural network approach to estimate reference evapotranspiration using limited climate data. *Neural Comput. Applic.* **2022**, *34*, 4013–4032. [[CrossRef](#)]
25. Xing, L.; Cui, N.; Guo, L.; Du, T.; Gong, D.; Zhan, C.; Zhao, L.; Wu, Z. Estimating daily reference evapotranspiration using a novel hybrid deep learning model. *J. Hydrol.* **2022**, *614*, 128567. [[CrossRef](#)]
26. Ferreira, L.B.; da Cunha, F.F. Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Comput. Electron. Agric.* **2020**, *178*, 105728. [[CrossRef](#)]
27. Luo, Y.; Chang, X.; Peng, S.; Khan, S.; Wang, W.; Zheng, Q.; Cai, X. Short-term forecasting of daily reference evapotranspiration using the Hargreaves–Samani model and temperature forecasts. *Agric. Water Manag.* **2014**, *136*, 42–51. [[CrossRef](#)]
28. Karbasi, M. Forecasting of multi-step ahead reference evapotranspiration using wavelet-Gaussian process regression model. *Water Resour. Manag.* **2018**, *32*, 1035–1052. [[CrossRef](#)]
29. Perera, K.C.; Western, A.W.; Nawarathna, B.; George, B. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. *Agric. Forest Meteorol.* **2014**, *194*, 50–63. [[CrossRef](#)]
30. Roy, D.K.; Sarkar, T.K.; Kamar, S.S.A.; Goswami, T.; Muktadir, M.A.; Al-Ghobari, H.M.; Alataway, A.; Dewidar, A.Z.; El-Shafei, A.A.; Mattar, M.A. Daily prediction and multi-step forward forecasting of reference evapotranspiration using LSTM and Bi-LSTM models. *Agronomy* **2022**, *12*, 594. [[CrossRef](#)]
31. Yin, J.; Deng, Z.; Ines, A.V.M.; Wu, J.; Rasu, E. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agric. Water Manag.* **2020**, *242*, 106386. [[CrossRef](#)]
32. Narapusetty, B.; DelSole, T.; Tippet, M.K. Optimal estimation of the climatological mean. *J. Clim.* **2009**, *22*, 4845–4859. [[CrossRef](#)]
33. McMahon, T.A.; Peel, M.C.; Lowe, L.; Srikanthan, R.; McVicar, T.R. Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: A pragmatic synthesis. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1331–1363. [[CrossRef](#)]
34. Zotarelli, L.; Dukes, M.D.; Romero, C.C.; Migliaccio, K.W.; Morgan, K.T. *Step by Step Calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method)*; Institute of Food and Agricultural Sciences, University of Florida: Gainesville, FL, USA, 2010; Volume AE459, pp. 1–10.
35. Vicente-Serrano, S.M.; Azorin-Molina, C.; Sanchez-Lorenzo, A.; Revuelto, J.; López-Moreno, J.I.; González-Hidalgo, J.C.; Moran-Tejeda, E.; Espejo, F. Reference evapotranspiration variability and trends in Spain, 1961–2011. *Glob. Planet. Change* **2014**, *121*, 26–40. [[CrossRef](#)]
36. Hargreaves, G.H.; Samani, Z.A. Reference crop evapotranspiration from temperature. *Appl. Eng. Agric.* **1985**, *1*, 96–99. [[CrossRef](#)]
37. De Bruin, H.A.R.; Lablans, W.N. Reference crop evapotranspiration determined with a modified Makkink equation. *Hydrol. Process.* **1998**, *12*, 1053–1062. [[CrossRef](#)]
38. Valiantzas, J.D. Simplified forms for the standardized FAO-56 Penman-Monteith reference evapotranspiration using limited weather data. *J. Hydrol.* **2013**, *505*, 13–23. [[CrossRef](#)]
39. Kuo, C.-E.; Chen, G.-T. Automatic sleep staging based on a hybrid stacked LSTM neural network: Verification using large-scale dataset. *IEEE Access* **2020**, *8*, 111837–111849. [[CrossRef](#)]
40. Lin, Y.-S.; Fang, S.-L.; Kang, L.; Chen, C.-C.; Yao, M.-H.; Kuo, B.-J. Combining recurrent neural network and sigmoid growth models for short-term temperature forecasting and tomato growth prediction in a plastic greenhouse. *Horticulturae* **2024**, *10*, 230. [[CrossRef](#)]
41. Wunsch, A.; Liesch, T.; Broda, S. Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1671–1687. [[CrossRef](#)]
42. Rahimi Khoob, A. Comparative study of Hargreaves’s and artificial neural network’s methodologies in estimating reference evapotranspiration in a semiarid environment. *Irrig. Sci.* **2008**, *26*, 253–259. [[CrossRef](#)]
43. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [[CrossRef](#)] [[PubMed](#)]
44. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]

45. Córdova, M.; Carrillo-Rojas, G.; Crespo, P.; Wilcox, B.; Célleri, R. Evaluation of the Penman-Monteith (FAO 56 PM) method for calculating reference evapotranspiration using limited data. *Mt. Res. Dev.* **2015**, *35*, 230–239. [[CrossRef](#)]
46. Fisher, J.B.; Terry, A.; DeBiase, A.; Qi1, Y.; Xu, M.; Allen, H. Evapotranspiration models compared on a Sierra Nevada forest ecosystem. *Environ. Model. Softw.* **2005**, *20*, 783–796. [[CrossRef](#)]
47. Makwana, J.J.; Tiwari, M.K.; Deora, B.S. Development and comparison of artificial intelligence models for estimating daily reference evapotranspiration from limited input variables. *Smart Agric. Technol.* **2023**, *3*, 100115. [[CrossRef](#)]
48. Traore, S.; Wang, Y.-M.; Kerh, T. Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. *Agric. Water Manag.* **2010**, *97*, 707–714. [[CrossRef](#)]
49. Jain, S.K.; Nayak, P.C.; Sudheer, K.P. Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. *Hydrol. Process. An Int. J.* **2008**, *22*, 2225–2234. [[CrossRef](#)]
50. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [[CrossRef](#)]
51. Wu, T.; Zhang, W.; Jiao, X.; Guo, W.; Hamoud, Y.A. Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Comput. Electron. Agric.* **2021**, *184*, 106039. [[CrossRef](#)]
52. Bellido-Jimenez, J.A.; Estevez, J.; Garcia-Marin, A.P. New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain. *Agric. Water Manag.* **2021**, *245*, 106558. [[CrossRef](#)]
53. Bellido-Jimenez, J.A.; Estevez, J.; Garcia-Marin, A.P. A regional machine learning method to outperform temperature-based reference evapotranspiration estimations in Southern Spain. *Agric. Water Manag.* **2022**, *274*, 107955. [[CrossRef](#)]
54. Su, Y.-C.; Kuo, B.-J. Risk assessment of rice damage due to heavy rain in Taiwan. *Agriculture* **2023**, *13*, 630. [[CrossRef](#)]
55. Esparza-Gómez, J.M.; Luque-Vega, L.F.; Guerrero-Osuna, H.A.; Carrasco-Navarro, R.; García-Vázquez, F.; Mata-Romero, M.E.; Olvera-Olvera, C.A.; Carlos-Mancilla, M.A.; Solís-Sánchez, L.O. Long short-term memory recurrent neural network and extreme gradient boosting algorithms applied in a greenhouse’s internal temperature prediction. *Appl. Sci.* **2023**, *13*, 12341. [[CrossRef](#)]
56. Shiri, J.; Nazemi, A.H.; Sadraddini, A.A.; Landeras, G.; Kisi, O.; Fard, A.F.; Marti, P. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Comput. Electron. Agric.* **2014**, *108*, 230–241. [[CrossRef](#)]
57. Jato-Espino, D.; Charlesworth, S.M.; Perales-Momparler, S.; Andrés-Doménech, I. Prediction of evapotranspiration in a Mediterranean region using basic meteorological variables. *J. Hydrol. Eng.* **2017**, *22*, 04016064. [[CrossRef](#)]
58. Hsu, H.H.; Chen, C.T. Observed and projected climate change in Taiwan. *Meteorol. Atmos. Phys.* **2002**, *79*, 87–104. [[CrossRef](#)]
59. Yu, P.S.; Yang, T.C.; Wu, C.K. Effects of climate change on evapotranspiration from paddy fields in southern Taiwan. *Clim. Change* **2002**, *54*, 165–179. [[CrossRef](#)]
60. de la Casa, A.C.; Ovando, G.G. Variation of reference evapotranspiration in the central region of Argentina between 1941 and 2010. *J. Hydrol. Reg. Stud.* **2016**, *5*, 66–79. [[CrossRef](#)]
61. Fan, J.; Wu, L.; Zhang, F.; Xiang, Y.; Zheng, J. Climate change effects on reference crop evapotranspiration across different climatic zones of China during 1956–2015. *J. Hydrol.* **2016**, *542*, 923–937. [[CrossRef](#)]
62. Łabędzki, L.; Bąk, B.; Smarzyńska, K. Spatio-temporal variability and trends of Penman-Monteith reference evapotranspiration (FAO-56) in 1971–2010 under climatic conditions of Poland. *Pol. J. Environ. Stud.* **2014**, *23*, 2083–2091. [[CrossRef](#)] [[PubMed](#)]
63. Maček, U.; Bezak, N.; Šraj, M. Reference evapotranspiration changes in Slovenia, Europe. *Agric. For. Meteorol.* **2018**, *260*, 183–192. [[CrossRef](#)]
64. Mojid, M.A.; Rannu, R.P.; Karim, N.N. Climate change impacts on reference crop evapotranspiration in North-West hydrological region of Bangladesh. *Int. J. Climatol.* **2015**, *35*, 4041–4046. [[CrossRef](#)]
65. Shadmani, M.; Marofi, S.; Roknian, M. Trend analysis in reference evapotranspiration using Mann-Kendall and Spearman’s Rho tests in arid regions of Iran. *Water Resour. Manag.* **2012**, *26*, 211–224. [[CrossRef](#)]
66. Tabari, H.; Marofi, S.; Aeini, A.; Talaee, P.H.; Mohammadi, K. Trend analysis of reference evapotranspiration in the western half of Iran. *Agric. For. Meteorol.* **2011**, *151*, 128–136. [[CrossRef](#)]
67. Zhang, D.; Liu, X.; Hong, H. Assessing the effect of climate change on reference evapotranspiration in China. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 1871–1881. [[CrossRef](#)]
68. Chia, M.Y.; Huang, Y.F.; Koo, C.H.; Fung, K.F. Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—A review. *Agronomy* **2020**, *10*, 101. [[CrossRef](#)]
69. Fang, S.-L.; Chang, T.-J.; Tu, Y.-K.; Chen, H.-W.; Yao, M.-H.; Kuo, B.-J. Plant-response-based control strategy for irrigation and environmental controls for greenhouse tomato seedling cultivation. *Agriculture* **2022**, *12*, 633. [[CrossRef](#)]
70. Incrocci, L.; Thompson, R.B.; Fernandez-Fernandez, M.D.; De Pascale, S.; Pardossi, A.; Stanghellini, C.; Roupshael, Y.; Gallardo, M. Irrigation management of European greenhouse vegetable crops. *Agric. Water Manag.* **2020**, *242*, 106393. [[CrossRef](#)]
71. Fernández, M.D.; Bonachela, S.; Orgaz, F.; Thompson, R.; López, J.C.; Granados, M.R.; Gallardo, M.; Fereres, E. Measurement and estimation of plastic greenhouse reference evapotranspiration in a Mediterranean climate. *Irrig. Sci.* **2010**, *28*, 497–509. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.