atmosphere

MDPI

*Article*

# Performance-Based Comparative Forecasting of Near-Future Evapotranspiration Using Statistical, Machine-Learning and Deep Learning Methods: A Case Study of Lake Burdur, Türkiye

**Muzaffer Göztaş [1,*], Nida Oruç Ünal [2], Doğan Yıldız [1] and Dursun Yıldız [3]**

[1] Department of Statistics, Faculty of Arts and Sciences, Yıldız Technical University, Istanbul 34000, Türkiye
[2] Department of Management Information Systems, Faculty of Economics, Administrative and Social Sciences, Istanbul Topkapi University, Istanbul 34000, Türkiye
[3] Hidropolitik Akademi, Istanbul 34000, Türkiye
* Correspondence: muzaffer.goztas@yildiz.edu.tr

**Abstract**

In this study, daily reference evapotranspiration ($ET_0$) values for the period 2025–2030 for Lake Burdur, located in the Mediterranean climate zone and within the Burdur closed basin, were estimated using nested architecture focused on high accuracy. The $ET_0$ target corresponds to the FAO-56 Penman–Monteith reference evapotranspiration variable provided by the Open-Meteo Historical Weather API, and it is treated throughout as a standardized measure of atmospheric evaporative demand rather than as actual lake-surface evaporation or basin water loss. For this purpose, daily mean air temperature, relative humidity, shortwave surface radiation, and evapotranspiration data for the period 1984–2024 were obtained from the Open-Meteo platform. In the first stage of the study (Model 1), separate SARIMAX (statistical), XGBoost (machine learning), and LSTM (deep learning) models were applied for temperature, relative humidity, and radiation series; the model with the highest validation mean for each variable was selected. Accordingly, LSTM (Mean $R^2$ = 0.967) was determined to be the most successful model for temperature, SARIMA(X) (Mean $R^2$ = 0.812) for relative humidity, and XGBoost (Mean $R^2$ = 0.845) for the radiation variable, which is non-linear, has strong autocorrelation, and exhibits distinct seasonality. In the second stage (Model 2), these best climate predictions were used as independent variables for evapotranspiration, and LSTM provided the highest success for evapotranspiration (Mean $R^2$ = 0.941). Trend analyses revealed that the increase in temperature and evapotranspiration and the decrease in relative humidity observed in the past period will continue in the near future. The uncertainty analysis conducted using the Monte Carlo/resampling approach on historical data showed that the 95% prediction intervals largely protected the upward trend in evapotranspiration against random fluctuations. These intervals reflect residual-based uncertainty under the fitted model rather than the full predictive uncertainty of future basin evapotranspiration. The findings indicate that designing model selection appropriate to the structure of the variables within a nested prediction framework significantly improves forecast accuracy and can provide a viable decision support input for sustainable water management in Mediterranean basins experiencing water scarcity.

**Keywords:** evapotranspiration; reference evapotranspiration; time series forecasting; machine learning; deep learning; Lake Burdur; near-future forecasting

Check for updates

Academic Editors: Nikolaos Mihalopoulos, Aristomenis Karageorgis, Anastasios I. Stamou, Manolis Plionis and Dimitrios Melas

Received: 11 June 2026
Revised: 24 June 2026
Accepted: 30 June 2026
Published: 8 July 2026

**Copyright:** © 2026 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.

## 1. Introduction

Climate change is increasingly transforming hydrological regimes by affecting water resources that are already under pressure, particularly in semi-arid and Mediterranean basins [1]. Rising air temperatures, changes in relative humidity, and shifts in radiation patterns directly affect evapotranspiration ($ET_0$), a fundamental component of both the terrestrial water budget and surface energy balance [2,3]. Therefore, accurate estimation of evapotranspiration is critical for irrigation planning, drought risk assessment, and long-term water allocation planning, especially in regions with limited data and rapidly declining water levels [4]. Lakes such as Lake Burdur, located in southwestern Türkiye and forming a closed basin system, are extremely sensitive to these changes because increased evaporation losses cannot be compensated for by outflow. This situation can accelerate lake shrinkage, increased salinity, and related ecological and socio-economic impacts.

Historically, evapotranspiration has been estimated using a wide range of methods, including evaporation pan measurements, water budget approaches, physically based energy balance formulations, and mass transfer equations. In areas where detailed meteorological observations are available, the FAO-56 (Food and Agriculture Organization Irrigation and Drainage Paper No. 56) Penman–Monteith equation has become the de facto standard, while in data-limited regions, temperature-based empirical formulas and evaporation pan methods remain prevalent due to their simplicity and low computational costs [5,6]. However, these classical approaches often fail to adequately represent the non-linear interactions between multiple atmospheric variables, can produce biased results under non-stationary climate conditions, and can be highly sensitive to missing/poor-quality inputs [7,8]. This situation has increased interest over time in data-driven techniques that can approximately learn the complex relationships between climate variables and evapotranspiration without requiring strong assumptions about functional form [9–12].

Evapotranspiration estimates play a vital role in water resource management and optimization. In many studies, accurate estimation of reference or actual evapotranspiration has enabled more precise determination of irrigation water requirements in terms of timing and quantity, leading to concrete gains such as water savings, reduced energy costs, and increased yields. Studies conducted in different regions in recent years have demonstrated that satellite-based evapotranspiration data enables the monitoring of water use in agricultural fields at the field scale, providing tangible benefits such as improving water budgets, facilitating regulatory compliance, and achieving savings. For example, the OpenET platform developed by Melton et al. (2022) in the United States provides 30 m resolution evapotranspiration maps for agricultural areas in western states, enabling farmers and water authorities to accurately measure land-based water consumption [13]. This system has been adopted as an alternative to expensive and error-prone measurement devices in the legal reporting of water use in California's Sacramento–San Joaquin Delta, thereby reducing regulatory compliance costs while increasing data consistency. Similarly, in Africa, Chukalla et al. (2022) applied an evapotranspiration-based irrigation performance assessment at a large-scale sugarcane operation in Mozambique, comparing the efficiency of different irrigation methods [14]. This study demonstrated the potential for improving water use efficiency in irrigation schemes through the low-cost use of satellite data and shed light on water management decisions in similar large-scale operations.

In agricultural production, field applications of evapotranspiration estimate also provide significant benefits in terms of irrigation planning and crop yield. A decision support system project conducted with farmer participation in the Urmia Lake basin in Iran achieved significant water savings by implementing precise irrigation planning using real-time evapotranspiration calculations. Amini et al. (2025) demonstrated that, based on eight farms, optimization based on evapotranspiration resulted in 41% less water use with drip irrigation and 14% less with sprinkler irrigation, while also increasing the contribution of water to crop yield [15]. This allowed farmers to conserve the Urmia basin's limited water resources by consuming less water while maintaining the same yield. Similarly, in a field trial conducted in California, USA, Cahn et al. (2025) studied cauliflower production using evapotranspiration-based irrigation scheduling [16]. The results revealed that producers in the region achieved maximum yield by applying approximately 30% less irrigation water than calculated in traditional practices, meaning that additional irrigation did not increase yield. This study demonstrated that irrigation corresponding to 100% evapotranspiration requirements can maintain product quality and quantity while reducing farmers' unnecessary water use and energy costs. Thus, irrigation timing supported by evapotranspiration estimates increases agricultural water efficiency and reduces groundwater consumption and energy costs associated with over-irrigation.

These international studies, conducted in different climate zones (semi-arid, humid, Mediterranean, monsoon, etc.), at different temporal resolutions and using various model families (statistical, tree-based ML, artificial neural networks, hybrid/ensemble structures) for reference or actual evapotranspiration estimation, have become a critical tool for irrigation planning, drought preparedness, and watershed-scale water budget management. However, a significant portion of existing studies operate under a single-stage structure (e.g., direct evapotranspiration estimation from meteorological inputs) and do not explicitly model the hierarchical and temporal dependencies between independent climate variables and evapotranspiration as a multi-stage nested estimation chain. In this context, this study conducted on Lake Burdur establishes (i) a nested two-stage prediction architecture where independent climate variables (temperature, radiation, relative humidity) are first predicted separately in the first model using SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables), XGBoost (Extreme Gradient Boosting), and LSTM (Long Short-Term Memory), and then (ii) these predictions are fed into the final evapotranspiration prediction as input in the second model, thereby establishing a nested two-stage prediction architecture that aims to contribute something unique to the literature. It should be noted that chained, hierarchical, and hybrid climate–hydrology forecasting frameworks already exist in the literature, so the two-stage idea is not in itself methodologically new. Rather than claiming a fundamentally novel architecture, the contribution of this study is more specific: (a) the algorithm choice is tailored to the statistical structure of each driver variable (SARIMAX, XGBoost, and LSTM are compared per variable and the best is retained), (b) the whole chain is embedded in a 37-fold rolling-origin cross-validation that preserves temporal order and prevents information leakage at both stages, and (c) the design is applied to a data-scarce Mediterranean closed basin (Lake Burdur) for which no comparable near-future $ET_0$ forecast has been reported.

## 2. Materials and Methods

### *2.1. Study Area*

The study area is formed by Lake Burdur, located in the Burdur Closed Basin in southwestern Türkiye, represented by approximately 37.727779° N latitude and 30.169121° E longitude. This closed basin structure, which has no internal drainage, places the lake at the focal point of the regional hydrological cycle and makes it highly sensitive to changes in surface/groundwater components. The lake is located in a semi-arid Mediterranean climate zone, within an environmental matrix characterized by intensive agricultural activities and water use for irrigation purposes. It is of strategic importance at the regional scale, both in terms of its ecological functions and water resources management. Figure 1 presents satellite images (1991, 2005, and 2019) showing its location within Türkiye on a global scale, its basin location within the country's borders, and the lake's surroundings, visually illustrating the Burdur Lake's significant area shrinkage and shoreline retreat over time. This situation forms the basis of the study, making the lake a suitable natural laboratory for examining lake-atmosphere interactions and evaporation/evapotranspiration dynamics.

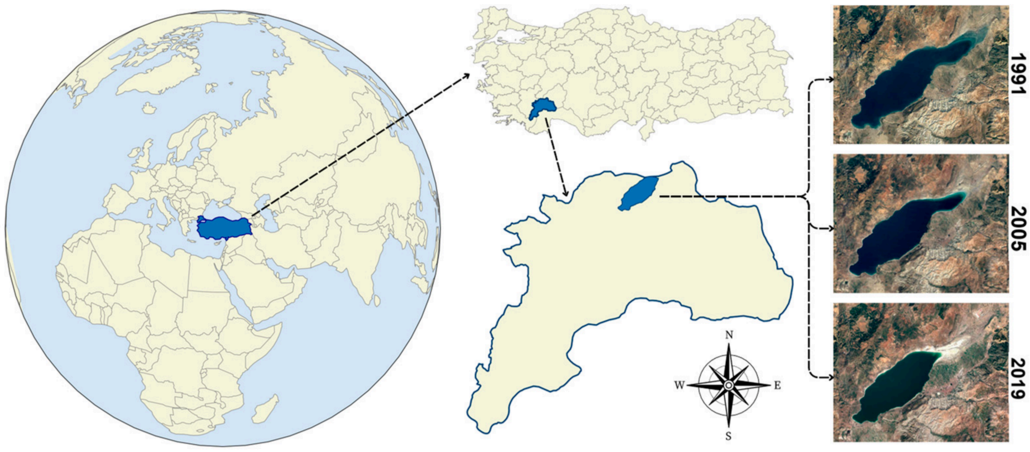


**Figure 1.** Location of the Burdur Lake study area and changes in surface area over time.

### *2.2. Data Processing*

First, the relationships between the dependent variable, evapotranspiration, and the explanatory variables (temperature, shortwave radiation, and relative humidity) were examined. Scatter plots and appropriate regression lines show that evapotranspiration increases significantly as temperature and radiation increase, whereas it decreases as relative humidity increases (see Figure 2). The target variable used throughout this study is daily reference evapotranspiration ($ET_0$), retrieved from the Open-Meteo Historical Weather API. In this archive, $ET_0$ is computed with the FAO-56 Penman–Monteith equation over a standardized short reference grass surface, using hourly reanalysis fields (2 m air temperature, relative humidity, shortwave radiation, and wind speed) that the platform redistributes from the ERA5/ERA5-Land reanalysis produced by ECMWF (European Centre for Medium-Range Weather Forecasts). Accordingly, $ET_0$ here represents a standardized atmospheric evaporative-demand quantity and is distinct from potential evapotranspiration, actual (terrestrial) evapotranspiration, and open-water lake evaporation, which depend additionally on soil moisture, vegetation state, land cover, surface resistance, and lake-surface energy exchange. The explanatory meteorological variables (air temperature, relative humidity, and downward shortwave surface radiation) were obtained from the same gridded product for the cell nearest the lake (approximately 37.73° N, 30.17° E) at native daily resolution; because all variables originate from a single reanalysis-derived grid point, the analysis should be understood as a forecast of a reanalysis-based $ET_0$ product, as discussed in the Limitations section. Since deviation from normal distribution was confirmed by Kolmogorov-Smirnov tests ($p$-value < 0.001 for all variables), correlations between variables were assessed using Spearman's rho ($\rho$) and Kendall's tau ($\tau$) coefficients [17,18]. The correlation analysis results indicate that the strongest positive correlation is between evapotranspiration and radiation ($\rho$ = 0.97, $\tau$ = 0.84), followed by temperature ($\rho$ = 0.90, $\tau$ = 0.70), and that relative humidity showed a strong but negative relationship with evapotranspiration ($\rho = -0.87$, $\tau = -0.68$). (Detailed distribution and correlation matrix are presented in Appendix Figure A1).

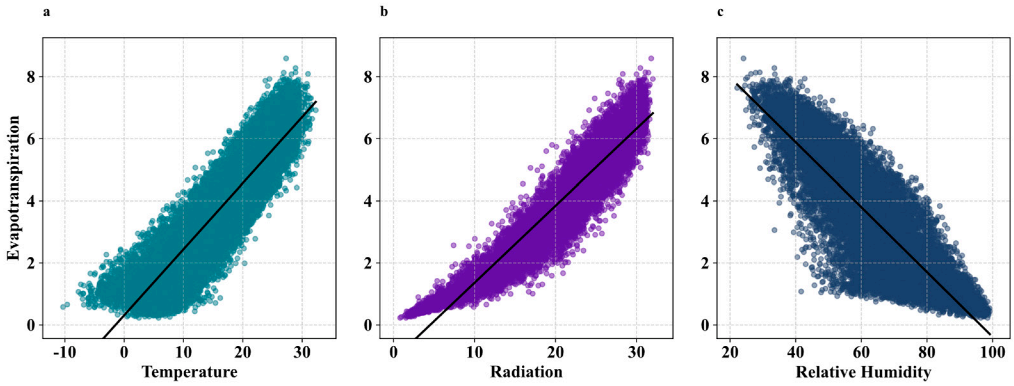


**Figure 2.** Scatterplots of daily evapotranspiration versus (**a**) temperature, (**b**) radiation and (**c**) relative humidity, with fitted linear regression lines indicating the direction and strength of the relationships.

To test whether stationarity, the fundamental assumption in modeling time series characteristics is satisfied. Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests were applied to the evapotranspiration, temperature, radiation, and relative humidity series [19,20]. The fact that the ADF test statistics for all variables are well below the critical value and the $p$-values are below 0.001 rejects the presence of a unit root, indicating that the series exhibit $I(0)$ behavior. Similarly, the test statistics obtained in both level and trend configurations in the KPSS test remained below the critical thresholds, with $p$-values at the 0.100 level, supporting the conclusion that the series are stationary around a constant level and/or deterministic trend. A combined interpretation of the ADF and KPSS results reveals that all variables have the same integration order ($I(0)$), thus indicating that multivariate time series models (SARIMA(X)) can be applied without the need for differencing ($d$ = 0) in Model 1, also consistent with recommended practice in recent hydroclimatic forecasting studies employing ARIMA/SARIMAX-type models [21] (detailed results are provided in Appendix Table A1).

$$y_t = \hat{S}_t + \hat{T}_t + \hat{R}_t \tag{1}$$

To quantitatively determine the seasonal structure in the series, STL decomposition was applied to each variable, and seasonal strength was calculated using an indicator based on the variance ratio. Accordingly, seasonal strength was found to be 0.941 for temperature, 0.926 for evapotranspiration, 0.873 for radiation, and 0.780 for relative humidity. According to Hyndman and Athanasopoulos, (2018) classification, values above 0.80 are interpreted as "very strong" seasonality, while the range 0.60–0.80 is interpreted as "strong" seasonality [22]. Therefore, it can be said that the first three variables exhibit a distinct and high-amplitude yearly cycle, while relative humidity, although relatively lower in amplitude, still contains a strong seasonal component. This finding indicated that seasonality should be represented directly within the model rather than removed from the data and formed the basis for using yearly periodic ($s$ = 365) seasonal parameters in the SARIMAX model (the seasonality parameter obtained from the STL decomposition is presented as $\hat{S}_t$ in Equation (1)). In Equation (1), $y_t$ denotes the observed value of a series at time $t$, $\hat{S}_t$ is the seasonal component, $\hat{T}_t$ is the trend component, and $\hat{R}_t$ is the remainder (irregular) component, all from the STL decomposition; $S_t$ and $T_t$ are subsequently represented inside the SARIMA(X) structure, while $\hat{R}_t$ is treated as the residual the model renders white-noise-like.

To identify long-term trends, the trend structure of four key meteorological variables was analyzed using the Mann–Kendall test, and the results were evaluated together with the trend component (denoted as $\hat{T}_t$ in Equation (1)) obtained from the STL decomposition.

Positive and statistically highly significant increases were observed in the temperature and evapotranspiration series (e.g., $Z_{Test}$ = 9.04, $p < 0.001$ for temperature; $Z_{Test}$ = 4.52, $p < 0.001$ for evapotranspiration), indicating a pronounced warming and accompanying increase in evapotranspiration during the study period. For the relative humidity series, a negative and strong decreasing trend was detected ($Z_{Test}$ = 9.02, $p < 0.001$), supporting the signal of regional aridification. No significant trend was detected in the radiation series ($p > 0.05$). Considering the $I(0)$ result of the ADF and KPSS tests, this structure indicates that the series are stationary but have a "trend-stationary" character containing a deterministic trend, in line with the KPSS framework [20,22]. Therefore, the $d = 0$ approach was maintained without applying differencing, while a linear time trend term was added to the SARIMAX structure (as depicted in Equations (2) and (3)), and the positive trends in temperature and evapotranspiration and the negative trend in relative humidity were represented in a clear and reproducible manner through the model parameters. (Detailed statistics regarding the Mann–Kendall results are summarized in Table 1).

$$y_t = \beta_0 + x_t^{\mathrm{T}}\beta + \sum\nolimits_{\mathrm{i}=1}^{P} \phi_i y_{t-1} + \sum\nolimits_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{2}$$

$$y_t = \beta_0 + \beta_1 t + x_t^{\mathrm{T}}\beta + \sum\nolimits_{=1}^{P} \phi_i y_{t-1} + \sum\nolimits_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{3}$$

**Table 1.** Mann–Kendall results for trend assumption of variables. The "Seasonal Strength" column reports the STL-based seasonal-strength index (0–1) for each variable, while the *p*-value, Z-test, and Slope columns summarize the Mann–Kendall trend results; the sign of the slope indicates trend direction.

| Variable | Seasonal Strength | *p*-Value [1] | $Z_{Test}$ | Slope |
|---|---|---|---|---|
| Evapotranspiration | 0.926 | <0.001 | 4.25 | 0.000015 |
| Temperature | 0.941 | <0.001 | 9.04 | 0.000143 |
| Radiation | 0.873 | 0.374 | −0.89 | −0.000013 |
| Relative Humidity | 0.780 | <0.001 | −9.02 | −0.000300 |

[1] Statistical significance evaluated at α = 0.05.

### *2.3. Hyperparameter Tuning for Algorithms*

At this stage, a nested modeling framework has been established to address the hierarchical dependencies of time series variables and the challenges of forward forecasting (see Figure 3). The workflow consists of two interconnected stages. In Model 1, the independent variables—temperature, radiation, and relative humidity—were forecasted separately for the 1984–2024 period using three alternative algorithms (SARIMAX, XGBoost, and LSTM). In Model 2, the climate predictions obtained from Model 1 were used as external explanatory variables to model future evapotranspiration dynamics, and SARIMAX, XGBoost, and LSTM were again compared under the same evaluation setup. In Model 1, evapotranspiration was not used as an independent variable at any stage, so all climate variables were modeled solely through their own autocorrelation structures.

In the hyperparameter optimization process, instead of a fixed training/test/validation split, a rolling-origin time series cross-validation with expanding windows was applied, which preserves the temporal order in both stages and prevents information leakage. Following a four-year warm-up window (W = 1.460 days), the model was trained on the interval $[t_0, t_k]$ at each step and validated only on the forward block $(t_k, t_k + H]$ for a one-year prediction horizon (H = 365 days). This process was repeated across 37 layers for each variable–algorithm–hyperparameter combination. The mean validation $R^2$ obtained across the layers was used as the primary selection criterion (in case of a tie, lower RMSE was preferred, and mean AIC and BIC values were preferred in the SARIMA(X) model), and the best-performing configuration for each algorithm was selected for use in subsequent comparisons and interpretations.

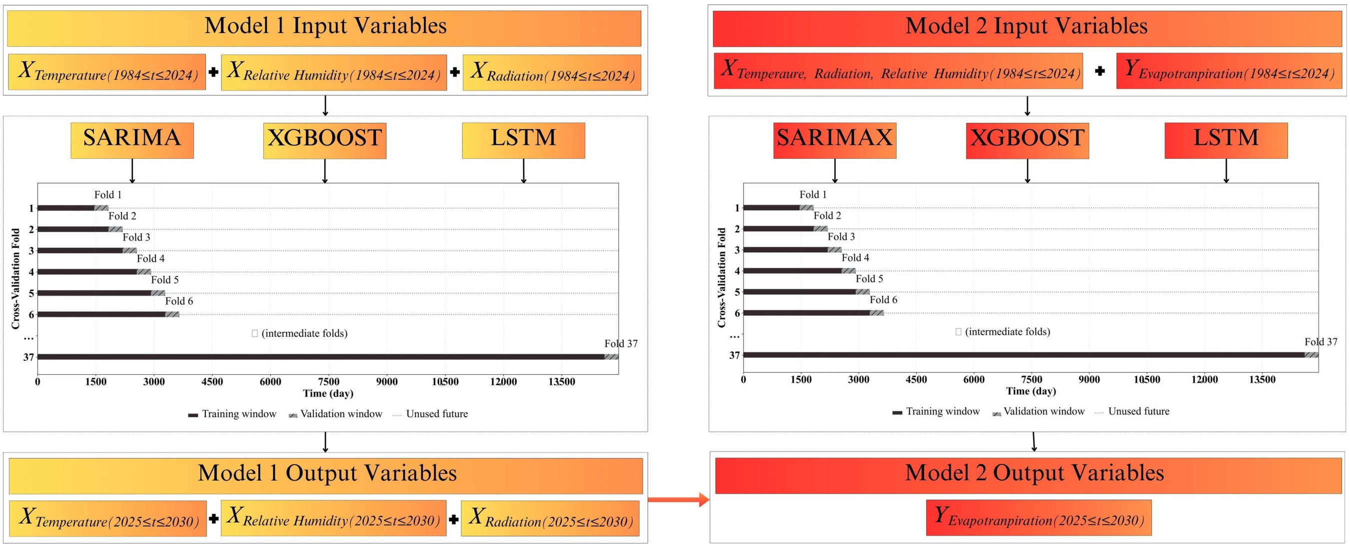


**Figure 3.** End-to-end schematic of the nested modelling workflow linking climate variable forecasting (Model 1) to evapotranspiration prediction (Model 2) using SARIMA, XGBoost, and LSTM.

### 2.3.1. SARIMA(X) Hyperparameter Tuning

Both Model 1 (temperature, radiation, relative humidity) and Model 2 (evapotranspiration) were constructed using SARIMA(X) structures, with $d = 0$ and $D = 0$ fixed to be consistent with stationarity findings. Confirmation that all series are $I(0)$ at the level using ADF and KPSS tests eliminated the need for differencing. Since the PACF (Partial Autocorrelation Function) outputs showed a distinct peak at the first 1–2 lags followed by rapid decay for all variables, low-order structures were preferred for both the seasonally adjusted and unadjusted parts (See details in Appendix Figure A2). Therefore, on the grid $p, q, P, Q \in \{0, 1, 2\}$, a total of $3^4 = 81$ candidate SARIMA(X) orders were screened for each series; the mean BIC was used as the primary criterion for selection, and AIC was used in case of equality (see Table 2).

**Table 2.** Mean AIC and BIC criterion range for best fit.

| Model | Predicted Variable | Criterion Range | | | | Best Fit | |
|---|---|---|---|---|---|---|---|
| | | AIC | | BIC | | AR | MA |
| | | Lower | Upper | Lower | Upper | **($p$, $P$)** | **($q$, $Q$)** |
| Model 1 | Temperature | 55,719.6 | 114,115.3 | 114,130.5 | 55,750.1 | (2, 1) | (2, 1) |
| Model 1 | Radiation | 79,113.5 | 117,182.0 | 77,144.0 | 117,197.0 | (2, 0) | (2, 1) |
| Model 1 | Relative Humidity | 105,267.1 | 151,374.1 | 105,305.2 | 151,389.3 | (2, 1) | (1, 0) |
| Model 2 | Evapotranspiration | −4863.3 | 5035.5 | −4809.6 | 5073.6 | (2, 1) | (2, 1) |

As a result of information criterion optimization, within Model 1, for temperature, $(p, d, q) = (2, 0, 2)$ and $(P, D, Q, s) = (1, 0, 1, 365)$; for radiation, (2, 0, 2) and (0, 0, 1, 365); and for relative humidity, (2, 0, 1) and (1, 0, 0, 365) were obtained as the most suitable orders. In Model 2, evapotranspiration was used as the dependent variable, and the temperature, radiation, and relative humidity estimates obtained from Model 1 were used as external variables; here too, the $(2, 0, 2) \times (1, 0, 1, 365)$ structure was the configuration that provided the lowest values in terms of both AIC and BIC (column "Best Fit" in Table 2). Due to the trend-stationary nature of the series, a linear deterministic trend term parameter $t$ was included in the evapotranspiration model, and the stationarity/invertibility constraints were relaxed in the parameter estimation (enforce_stationarity = False, enforce_invertibility = False).

The statistical adequacy of these selected low-order structures has been evaluated using sequential diagnostic tests applied to the residual processes. The Ljung-Box Q test $p$-values calculated in multiple lag windows (10, 20, 30, 60, and 365) for temperature, radiation, relative humidity, and evapotranspiration residuals remained above the 0.05 threshold value. The KPSS level and trend tests showed no conflict with the assumption of stationarity of the residuals, while the White and Breusch–Pagan tests showed that the assumption of homoscedasticity could not be rejected. In particular, the fact that the Ljung-Box $p$-values are above 0.05 at medium and long lag intervals [30, 365] supports that there is no significant residual autocorrelation after the model. Taken together, these findings indicate that the selected $p$, $P$, $q$, and $Q$ combinations produce white noise-like, constant variance residuals for both climate variables and evapotranspiration, and that the SARIMA(X) models offer a statistically robust structure in terms of forecasting performance (For detailed residual diagnosis results, see Appendix Table A2).

#### 2.3.2. XGBoost Hyperparameter Tuning

For each independent variable within Model 1, a systematic hyperparameter optimization was performed using the XGBoost algorithm. A grid search was conducted on four core hyperparameters (learning_rate, max_depth, subsample, colsample_bytree), with five candidate values defined for each, resulting in $5 \times 5 \times 5 \times 5 = 625$ distinct configurations being tested. All evaluations were performed using 37-layer expanding window rolling-origin cross-validation, which preserves the time order and prevents forward information leakage. The mean $R^2$ obtained across layers for each configuration was used as the primary selection criterion. Figure 4 jointly visualizes the $R^2$ surfaces for the learning_rate-max_depth pair, which determines XGBoost's learning dynamics, and the subsample-colsample_bytree pair, which controls stochastic regularization. Only the effect of the relevant pair is isolated in the panels, with other hyperparameters held constant.

As a result of XGBoost Hyperparameter Tuning process, a more moderate learning step and shallow tree structure were preferred to avoid excessive complexity for the temperature series. The combination of learning_rate = 0.005, max_depth = 3, subsample = 0.6, and colsample_bytree = 0.6 was selected (see Figure 4a,b). For radiation, $R^2$ values were seen to cluster in the 84–86% band, and the quartet learning_rate = 0.01, max_depth = 7, subsample = 0.8, colsample_bytree = 0.8 was determined to be the optimal structure that systematically improved validation performance (see Figure 4c,d). For the relative humidity variable, $R^2$ values in the range of 76–79% were obtained in the learning_rate–max_depth plane; learning_rate = 0.02 and max_depth = 3 stood out in terms of learning speed and generalization balance, while the combination of subsample = 0.9 and colsample_bytree = 1.0 minimized information loss and balanced variance with slight randomization (see Figure 4e,f). In Model 2, only the temperature, radiation, and relative humidity values of the same day were used as inputs for evapotranspiration; the relatively flat structure of the hyperparameter surfaces revealed that the model's sensitivity to these parameters was limited, and that a stable, low-variance solution was more rational than high accuracy. For this reason, values in the middle band were preferred for the evapotranspiration model (Model 2), with learning_rate = 0.03, max_depth = 5, subsample = 0.8, and colsample_bytree = 0.8 (see Figure 4g,h). In the study, n_estimators = 500 and random_state = 42 were adopted as a fixed structure that preserves generalization power (Figure 4).

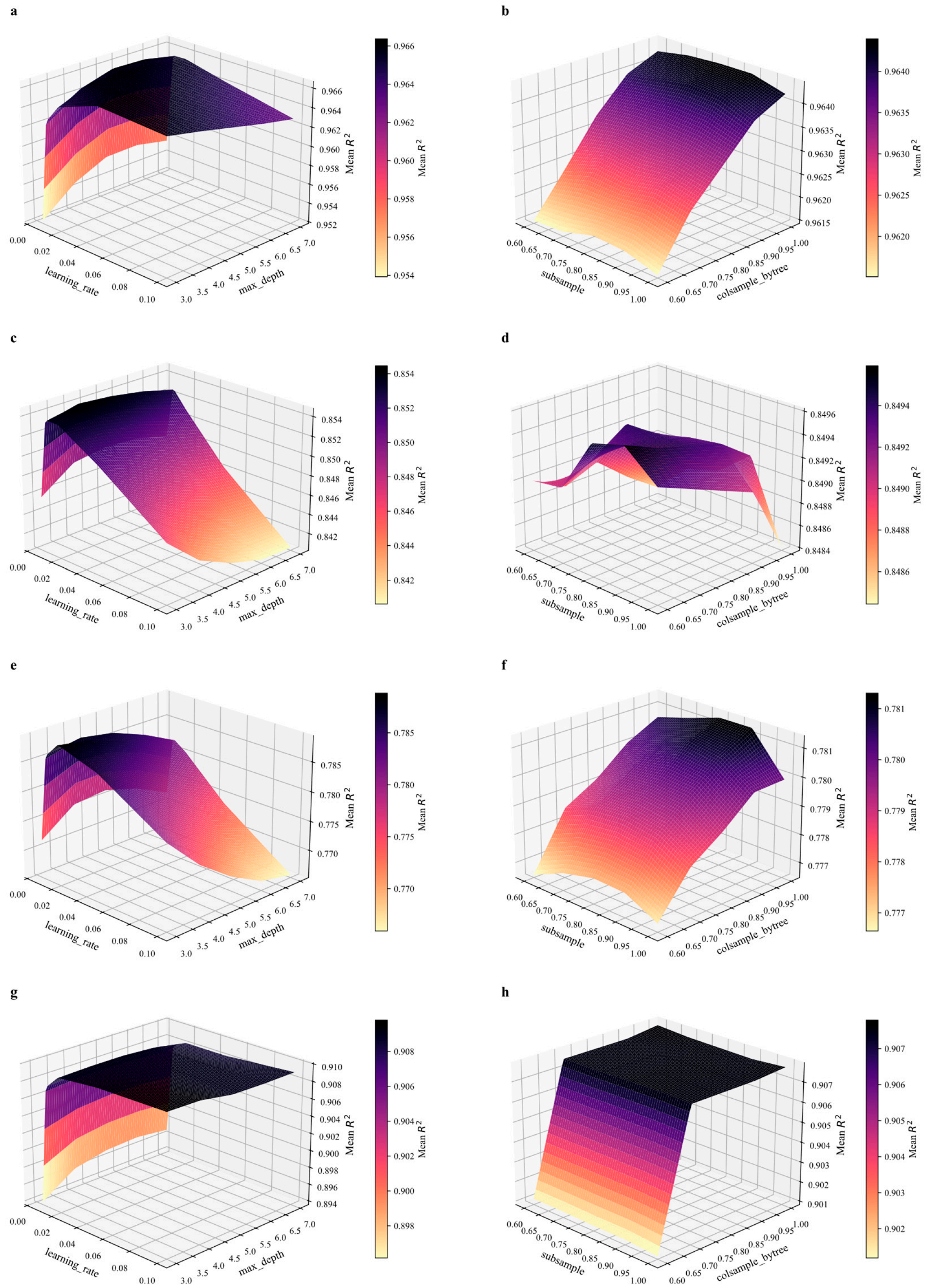

**Figure 4.** Mean $R^2$ levels for XGBoost hyperparameter sensitivity (learning_rate-max_depth and subsample-colsample_bytree); 37-fold Rolling-Origin Cross-Validation results by variable. Panels are organized by variable in two columns: the left panel of each pair (**a**,**c**,**e**,**g**) shows the mean validation $R^2$ surface over the learning_rate-max_depth plane, and the right panel (**b**,**d**,**f**,**h**) the subsample-colsample_bytree plane, for temperature (**a**,**b**), radiation (**c**,**d**), relative humidity (**e**,**f**), and evapotranspiration (**g**,**h**); other hyperparameters held constant.

### 2.3.3. LSTM Hyperparameter Tuning

The third prediction approach used in this study is Long Short-Term Memory (LSTM), a recurrent neural network architecture developed for modeling sequential dependencies. In classical RNNs, the "vanishing gradient" problem that arises during the learning of long-term dependencies weakens the flow of information over time, making it difficult to learn both short- and long-term patterns. Proposed by Hochreiter and Schmidhuber (1997), LSTM largely overcomes this problem thanks to three gates (input, forget, and output gates) that dynamically update the cell state; the input gate determines how much new information to take in, the forget gate determines which part of the past information to discard, and the output gate determines which information to transfer from the cell state to the next step [23]. In climate series containing strong autocorrelation and seasonality components, this structure provides a significant advantage over classical linear time series models due to its capacity to carry dependencies extending back hundreds of time steps. Recent studies focused on predicting hydro-meteorological processes such as evapotranspiration, runoff, and precipitation also show that LSTM-based models achieve high success, particularly in capturing seasonal patterns and climate change signals.

In the LSTM phase, variable-based hyperparameter searches were performed on the same basic architecture for both Model 1 (temperature, radiation, relative humidity) and Model 2 (evapotranspiration). Figure 5 schematically presents the optimized LSTM architecture used for each variable. AdamW was chosen as the optimization algorithm in all experiments, and the weight decay coefficient was fixed at 0.0001 without trying any additional combinations. This ensured L2-style regularization by penalizing excessive growth of layer weights and limited the risk of overfitting [24,25]. The loss function was defined as Mean Squared Error (MSE). The search space was created in the form of batch_size $\in$ {32, 64, 128, 256}, learning_rate $\in$ {0.0001, 0.0003, 0.001, 0.003}, and patience $\in$ {10, 15, 20}. Theoretically, 32 candidate configurations were randomly selected for each variable from $4 \times 4 \times 3 = 48$ possible structures (Randomized Search) and trained and compared. The evaluation was performed through 37-layer time series cross-validation according to the rolling-origin (expanding window) scheme presented in Figure 3. A 365-day windowing (input window) was used for the independent variables, training in each layer was performed only with data in the $[t_0, t_k]$ interval, and the forward block $[t_k, t_k + 365]$ was reserved for validation. The mean $R^2$ obtained across layers was used as the primary criterion, and the mean RMSE as the secondary criterion during the decision phase, when early stopping was triggered, training was terminated because the validation loss no longer improved.

Figure 6 summarizes the mean $R^2$-RMSE-epoch distributions and the best combinations obtained from 32 candidate structures for each variable. Within Model 1, the structure providing high accuracy and consistent generalization for the temperature series with $R^2 = 0.966$ and RMSE = 1.553 was determined to be the combination of batch_size = 256, learning_rate = 0.001, and patience = 15 (Figure 6a). The best result for radiation is $R^2 = 0.847$ and RMSE = 2.922. Despite the series containing more variable and short-term fluctuations, a stable learning regime was achieved with batch_size = 256, learning_rate = 0.003, and patience = 20 (Figure 6b). For relative humidity, the values $R^2 = 0.814$ and RMSE = 7.171 indicate that the combination of batch_size = 256, learning_rate = 0.0003, and patience = 15 provides balanced generalization in this series, which is more affected by noise and external influences (Figure 6c). In Model 2 for evapotranspiration prediction, the LSTM model was fed only with three external signals (temperature, radiation, relative humidity) from the same day and sin/cos calendar components representing seasonality. Past lags were not used as input. As a result of rolling-origin validation consisting of 37 layers on 32 structures selected from 48 candidate combinations, the hyperparameter set providing the highest $R^2$ and lowest RMSE together for evapotranspiration was determined to be batch_size = 256, learning_rate = 0.0003, and patience = 20. This combination is characterized by low variance across successive layers, limited fluctuation, and caution against overfitting (See Figure 6d).

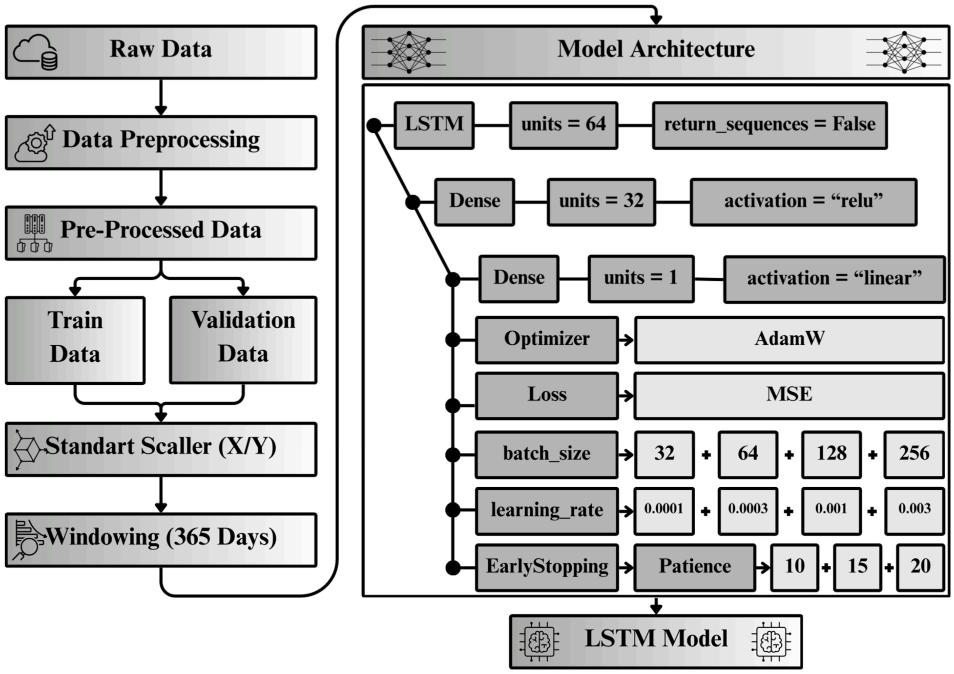


**Figure 5.** Optimized LSTM Architecture for each variable LSTM models.

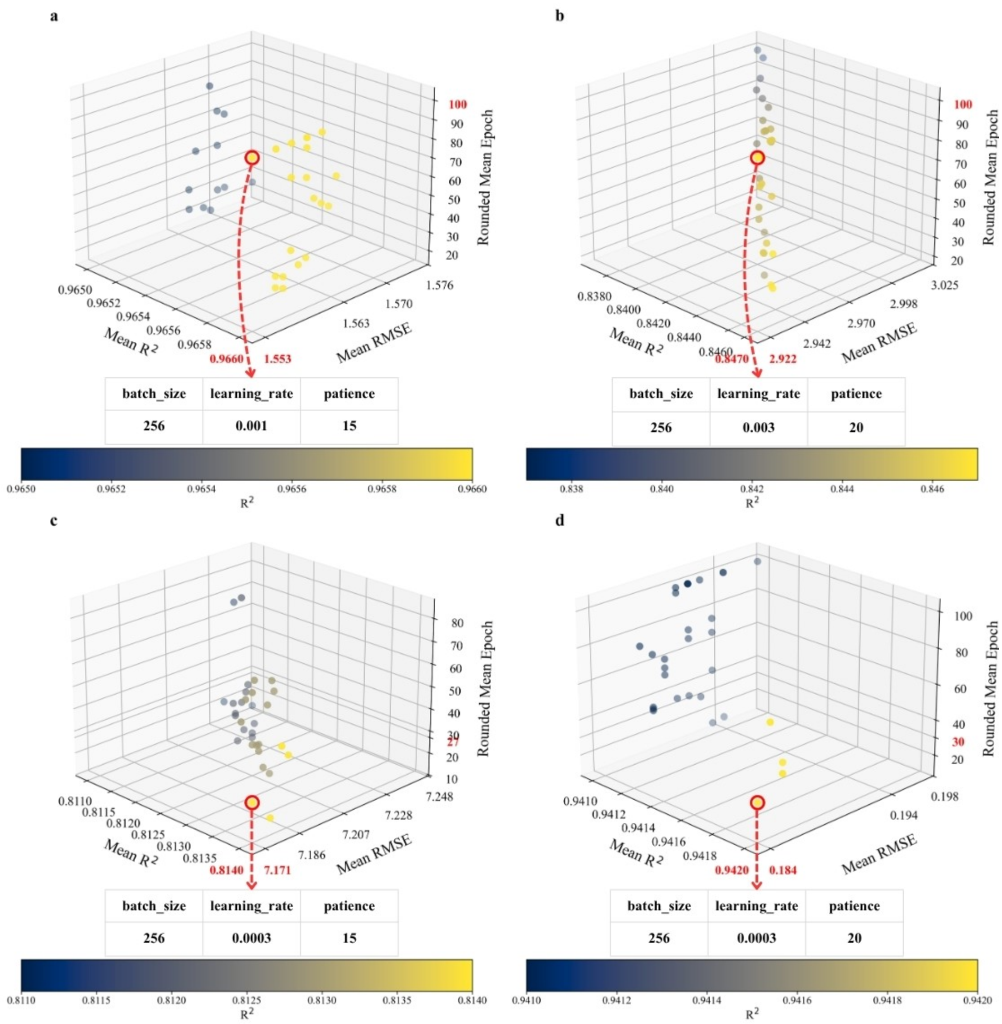


**Figure 6.** Hyperparameter tuning results for each LSTM model. Each subfigure summarizes mean $R^2$, RMSE, and epoch distributions across the 37 rolling-origin folds with the best combination for one variable: (**a**) temperature, (**b**) radiation, (**c**) relative humidity, and (**d**) evapotranspiration.

## 3. Results

### *3.1. Model Performance Comparison*

This section addresses the evaluation of model performance and the final model selection. First, the success metrics reported for each variable and algorithm are summarized in Table 3; maximization metrics include $R^2$, $R_{\alpha}^2$, Nash–Sutcliffe Efficiency (NSE), and Kling–Gupta Efficiency (KGE); while minimization criteria include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

**Table 3.** Model 1 and Model 2: Mean training and validation performance metrics for SARIMA(X), XGBoost, and LSTM models for temperature, radiation, relative humidity, and evapotranspiration under 37-layer rolling-origin cross-validation. (Note: Bold Values indicate the best-performing model for each variable based on the relevant performance criterion).

| Variable | Method | Step | Model 1 and 2 Success Criterion Means | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Maximization | | | | Minimization | | | |
| | | | $R^2$ | $R_{\alpha}^2$ | NSE | KGE | MSE | RMSE | MAE | MAPE |
| Temperature | SARIMA(X) | Train | 0.961 | 0.961 | 0.971 | 0.979 | 2.071 | 1.439 | 1.078 | 18.359 |
| | | Validation | 0.966 | 0.965 | 0.966 | 0.976 | 2.356 | 1.535 | 1.159 | 18.071 |
| | XGBoost | Train | 0.966 | 0.962 | 0.952 | 0.888 | 3.450 | 1.857 | 1.425 | 29.993 |
| | | Validation | 0.953 | 0.952 | 0.943 | 0.879 | 3.937 | 1.984 | 1.536 | 26.681 |
| | LSTM | Train | 0.967 | 0.961 | 0.971 | 0.981 | 2.412 | 1.553 | 1.067 | 17.002 |
| | | Validation | **0.967** | **0.965** | **0.967** | **0.980** | **2.336** | **1.528** | **1.150** | **16.241** |
| Radiation | SARIMA(X) | Train | 0.863 | 0.862 | 0.863 | 0.899 | 8.258 | 2.874 | 2.153 | 18.285 |
| | | Validation | 0.843 | 0.843 | 0.843 | 0.898 | 8.785 | 2.964 | 2.210 | 18.161 |
| | XGBoost | Train | 0.855 | 0.853 | 0.921 | 0.919 | 4.724 | 2.174 | 1.671 | 14.653 |
| | | Validation | **0.845** | **0.841** | **0.845** | **0.893** | **8.694** | **2.949** | **2.207** | **18.540** |
| | LSTM | Train | 0.847 | 0.846 | 0.867 | 0.895 | 8.538 | 2.922 | 2.105 | 17.920 |
| | | Validation | 0.827 | 0.827 | 0.847 | 0.893 | 8.560 | 2.919 | 2.076 | 17.993 |
| Relative Humidity | SARIMA(X) | Train | 0.800 | 0.800 | 0.800 | 0.851 | 55.023 | 7.418 | 5.742 | 9.328 |
| | | Validation | **0.812** | **0.807** | **0.817** | **0.856** | **51.990** | **7.210** | **5.653** | **10.255** |
| | XGBoost | Train | 0.785 | 0.785 | 0.815 | 0.852 | 50.859 | 7.132 | 5.562 | 9.057 |
| | | Validation | 0.793 | 0.788 | 0.813 | 0.852 | 51.715 | 7.191 | 5.641 | 10.338 |
| | LSTM | Train | 0.814 | 0.811 | 0.803 | 0.835 | 51.423 | 7.171 | 5.730 | 9.287 |
| | | Validation | 0.793 | 0.791 | 0.813 | 0.841 | 51.708 | 7.191 | 5.639 | 10.278 |
| Evapotranspiration | SARIMA(X) | Train | 0.932 | 0.932 | 0.982 | 0.987 | 0.074 | 0.272 | 0.213 | 10.889 |
| | | Validation | 0.937 | 0.937 | 0.977 | 0.945 | 0.093 | 0.305 | 0.234 | 9.323 |
| | XGBoost | Train | 0.918 | 0.915 | 0.995 | 0.996 | 0.020 | 0.141 | 0.097 | 3.183 |
| | | Validation | 0.920 | 0.920 | 0.990 | 0.967 | 0.042 | 0.205 | 0.138 | 4.091 |
| | LSTM | Train | 0.942 | 0.940 | 0.993 | 0.995 | 0.034 | 0.184 | 0.112 | 3.671 |
| | | Validation | **0.941** | **0.941** | **0.991** | **0.974** | **0.036** | **0.191** | **0.131** | **4.061** |

When Table 3 is evaluated holistically, it is observed that within the variable–algorithm–validation step combinations under Model 1, the highest accuracy values are provided by the LSTM model for temperature, the XGBoost model for radiation, and the SARIMA(X) model for relative humidity. For the temperature variable, in the LSTM validation step, approximately $R^2$ = 0.967, $R^2_\alpha$ = 0.964, NSE = 0.981, KGE = 0.980, MSE = 2.34, RMSE = 1.53, MAE = 1.15, and MAPE = 16.24. In the validation step for radiation, the XGBoost model produced the highest coefficient of determination with $R^2$ = 0.845 and $R^2_\alpha$ = 0.841, while the error metrics remained at similar levels. For relative humidity, the SARIMA(X) validation step stood out with approximately $R^2$ = 0.812, $R^2_\alpha$ = 0.807, NSE = 0.817, KGE = 0.856, and relatively low error values (e.g., RMSE = 7.21, MAE = 5.65). Overall, when examining both maximization criteria, such as $R^2/R^2_\alpha$, NSE, and KGE, and minimization criteria, such as MSE, RMSE, MAE, and MAPE, it can be said that the algorithm providing the highest accuracy for each variable is consistent with the algorithm providing the lowest error levels. Therefore, in the initial table-based evaluation, these three combinations (Temperature-LSTM, Radiation-XGBoost, Relative Humidity-SARIMA(X)) emerge as the preferred models for Model 1 in the initial table-based evaluation.

Under Model 2, the estimation model established for evapotranspiration uses only the prediction series produced by the best structures selected from Model 1 (Temperature-LSTM, Radiation-XGBoost, Relative Humidity-SARIMA(X)) as external variables. In the validation step for evapotranspiration, the LSTM model showed significantly superior performance compared to other algorithms, achieving approximately $R^2$ = 0.941, NSE = 0.991, KGE = 0.974, and very low error levels (MSE = 0.036, RMSE = 0.19, MAE = 0.13, MAPE = 4.06). Therefore, LSTM was chosen as the final prediction model for evapotranspiration, and the best predictions obtained from Model 1 were fed into this structure to produce forward-looking daily evapotranspiration predictions for the period 2025–2030. The agreement of the LSTM model with observed values was demonstrated by prediction-actual curves, which showed that it captured both the level and seasonal fluctuation dynamics to a large extent [26,27] (see Appendix Figure A3 for detailed prediction-actual graphs).

To test whether the algorithm–variable combinations selected in Model 1 and Model 2 are truly consistent and statistically distinguishable, paired sample *t*-tests were first applied to the fold-based $R^2$ values obtained from the 37-layer rolling-origin cross-validation (ROCV) process shown in Figure 6. In this context, for each variable (temperature, radiation, relative humidity, evapotranspiration), the train and validation steps of the same algorithm were first compared to investigate possible signs of overfitting, and then the $R^2$ level of the algorithm that best explained the relevant variable was tested pairwise against the results of the other algorithms. The condition of normal distribution of differences, which is the basic assumption of the paired sample *t*-test, was checked for each pair using the Shapiro–Wilk test. The fact that the *p*-values were above 0.05 for the vast majority of difference series indicated that the normal distribution assumption could not be rejected and, therefore, that the use of the parametric paired sample *t*-test was appropriate (the relevant Shapiro–Wilk *p*-values are presented in the heat maps in Appendix Figure A4). Thus, both the significance of the algorithm-internal train–validation differences and the statistical testing of performance differences between different algorithms predicting the same variable were ensured.

First, matched sample *t*-tests performed on the $R^2$ sequences obtained from the 37-layer rolling-origin cross-validation process were used to test the consistency of training and validation performance for each variable–algorithm combination (see Figure 7). The "within algorithms" cells highlighted in red in the table represent comparisons between Train–Validation pairs of the same algorithm. Across all panels (Figure 7a–d), *p*-values for Train–Validation matches remained systematically above the 0.05 threshold for SARIMA(X), XGBoost, and LSTM, meaning no statistically significant differences were observed in any algorithm–variable combination ($p > 0.05$). When evaluated together with the 37-layer $R^2$ curves shown in Figure 8, this result indicates that the training and validation steps performed at the same level in all models, thus showing no clear overfitting signal within the rolling-origin framework.

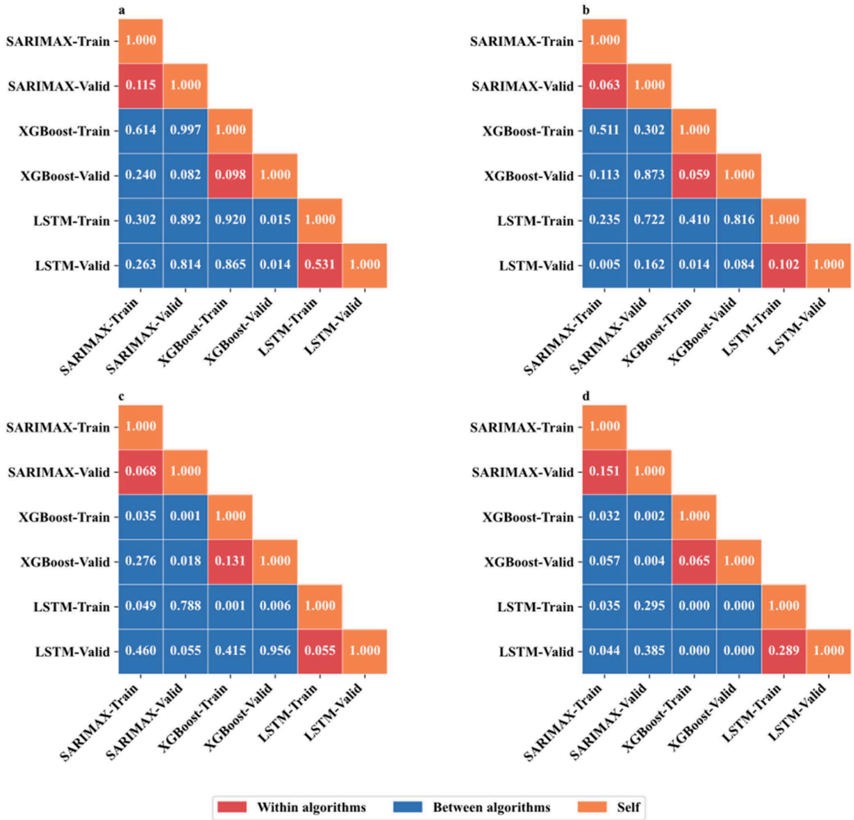


**Figure 7.** Paired sample *t*-test *p*-value matrices for within and between algorithm comparisons of $R^2$ across 37 rolling-origin folds for temperature (**a**), radiation (**b**), relative humidity (**c**), and evapotranspiration (**d**).

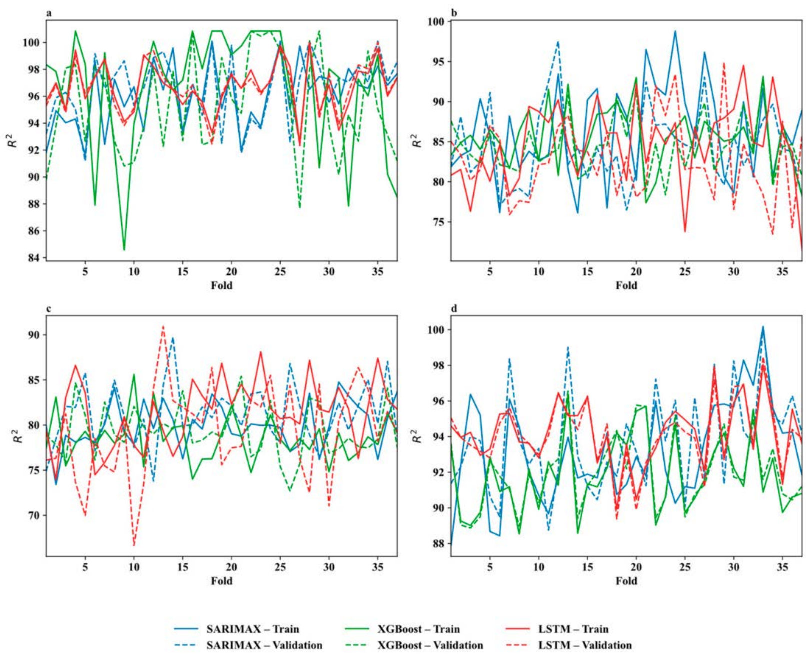


**Figure 8.** Fold-based $R^2$ curves for SARIMAX, XGBoost, and LSTM models for temperature (**a**), radiation (**b**), relative humidity (**c**), and evapotranspiration (**d**) in a 37-layer rolling-origin cross-validation scheme.

Comparisons between algorithms (blue cells in Figure 7) show that variable-based preferences are largely consistent with the mean criteria in Table 3, especially when interpreted through the validation step. (i) For temperature, LSTM stands out with the highest $R^2$ (0.967) in the validation step, but no significant difference is detected between the validation performances of LSTM and SARIMA(X) ($p > 0.05$; Figure 7a), while the difference is statistically significant in the LSTM-XGBoost validation comparison ($p < 0.01$). In this case, although LSTM has a statistically similar level of explanatory power to SARIMA(X), it has been the preferred algorithm for temperature due to its higher NSE and KGE values and lower MSE-RMSE-MAE-MAPE values (Table 3). (ii) In the validation step for the radiation variable, XGBoost stands out with the highest $R^2$ (0.845) and the lowest error metrics (especially MSE and RMSE). In Figure 7b, the *p*-values in the validation-based algorithm comparisons mostly remain above 0.05, indicating that the validation success of SARIMA(X), XGBoost, and LSTM is statistically very close to each other. In this case, the choice of XGBoost is based on its superiority in error-based metrics, in addition to small $R^2$ differences. (iii) For relative humidity, in the SARIMA(X) validation step in Table 3, it exhibits the best profile in terms of both $R^2$ (0.812) and error metrics, while in the validation comparisons in Figure 7c, some pairs show significant differences between SARIMA(X) and other algorithms ($p < 0.01$), which supports that SARIMA(X) is statistically more advantageous in explaining this variable. (iv) Finally, for evapotranspiration, Model 2, fed with the best predictions from Model 1, produced the highest $R^2$ (0.941) and the lowest error metrics in the LSTM validation step. Significant differences were obtained in the comparisons between the LSTM validation outputs and the SARIMA(X) and XGBoost validation outputs in Figure 7d ($p < 0.01$). Thus, LSTM was determined to be a statistically superior final prediction model for evapotranspiration based on both multiple success metrics and paired sample *t*-tests (Table 3, Figure 7d).

When conducting an overall evaluation, the algorithms selected for independent variables are not completely independent from other algorithms (each variable–algorithm pair involves a training and validation phase), but there are also cases where they are independent. However, the selection was made considering that the chosen algorithms yielded better mean results in other metrics.

### *3.2. 2025–2030 Forecasts*

In this section of the study, yearly means were used to facilitate interpretation when evaluating the estimated evapotranspiration values for the 2025–2030 period. Firstly, the temperature variable (see Figure 9a) shows a clear upward trend in yearly means during the actual period (blue dotted line). Short-term forecasts for the forecast period also continued this upward trend (red dots), and the slopes were largely parallel. Thus, the overall slope (green) was reinforced upward, ensuring consistent continuity between past patterns and forecasts. This indicates that temperature increases may continue to rise steadily in the coming years, parallel to previous years.

Looking at the radiation variable (see Figure 9b), the actual period exhibited a wavy but nearly horizontal behavior (blue dashed). Although a limited upward trend is forecasted for the forecast period (red dotted line), the overall slope (green) has remained near horizontal, indicating no significant break in the long-term total trend.

When examining the yearly means of the relative humidity variable in Figure 9c, a steady downward trend was observed in the actual period (blue dashed line). The forecast outputs have continued this downward trend (red dotted line), and the overall slope (green) has strengthened downward. This situation indicates that the structural pattern of the past has been consistently reflected in short-term forecasts.

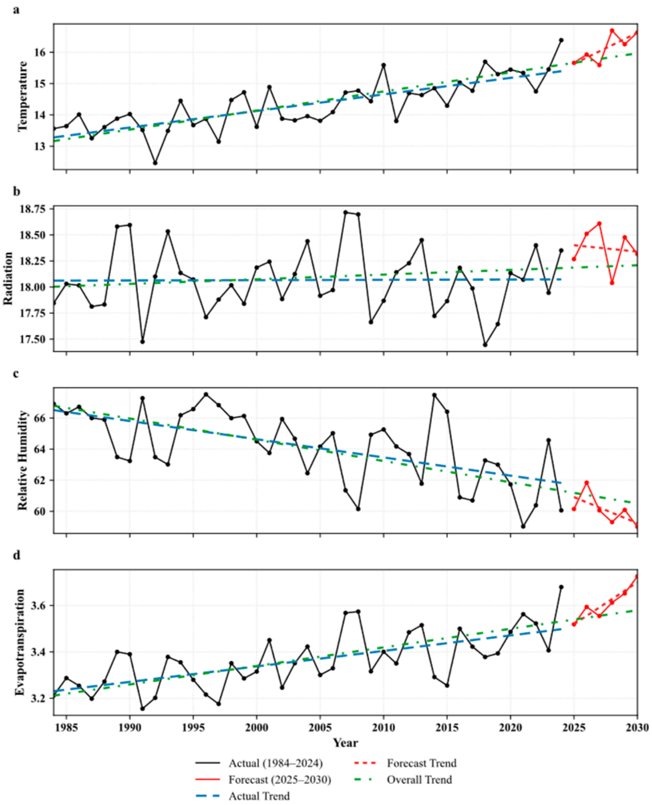


**Figure 9.** Yearly means, forecasts, and trend lines for all variables (yearly mean values and corresponding forecasts (2025–2030) for temperature (**a**), radiation (**b**), relative humidity (**c**), and evapotranspiration (**d**) are shown. For each variable, the historical period (1984–2024) is depicted in black, forecasts in red area. Linear trend lines are provided for the actual period (blue dashed), the forecast period (red dotted), and the overall period (green dash–dot)).

Figure 9d shows a slight but consistent upward trend in the evapotranspiration variable during the actual period (blue dotted line). Forecast projections continued this increase (red dotted line), and the overall slope (green) strengthened upward. Thus, it was concluded that the composite model output based on independent variable projections was consistent with past dynamics.

In summary, there is an upward trend for temperature and evapotranspiration, and a downward trend for relative humidity. For the radiation variable, no significant trend break was observed at the overall level. In this context, it can be stated that the forecast component is structurally consistent with actual trends.

Table 4 reports the linear slope coefficients, intercepts, and *p*-values of the Mann–Kendall trend test for the actual (1984–2024), forecast (2025–2030), and overall (1984–2030) periods, are interpreted consistently with the colour/line styles in Figure 9. The Mann–Kendall test examined whether there was a directional trend, while the sign and magnitude of the slope quantified the linear fit. Considering that the forecast period only includes 6 years of yearly means, it was assessed that the *p*-values were insignificant due to a decrease in power in the Mann–Kendall test, which was mostly due to the short number of observations in most variables; this situation should be seen as a statistical sensitivity problem arising from insufficient data.

**Table 4.** Linear trend slopes and intercepts of yearly means for each variables (actual term years [1984–2024], forecast term years [2025–2030], overall term years [1984–2030]. The colors and line styles of the "Process" column (actual: blue dashed, forecast: red dotted, overall: green dash-dot) are matched with the corresponding trend lines in Figure 9 for visual consistency and ease of comparison).

| Variable | Process | Slope | Intercept | $p$-Value [1] |
|---|---|---|---|---|
| Temperature | Actual | 0.05280 | −91.48 | <0.001 |
| | Forecast | 0.19864 | −386.63 | 0.259 |
| | Overall | 0.06089 | −107.64 | <0.001 |
| Radiation | Actual | 0.00029 | 17.49 | 0.727 |
| | Forecast | −0.01236 | 43.44 | 1.000 |
| | Overall | 0.00448 | 9.12 | 0.107 |
| Relative Humidity | Actual | −0.11707 | 298.79 | <0.001 |
| | Forecast | −0.33552 | 740.34 | 0.133 |
| | Overall | −0.13659 | 337.78 | <0.001 |
| Evapotranspiration | Actual | 0.00667 | −10.01 | <0.001 |
| | Forecast | 0.03621 | −69.81 | 0.024 |
| | Overall | 0.00799 | −12.65 | <0.001 |

[1] Statistical significance evaluated at $\alpha = 0.05$.

The slope of the temperature variable was positive and significant during the actual period ($p < 0.001$). Consistent with the blue dashed line in Figure 9a, an upward trend was observed throughout 1984–2024. During the forecast period, the slope was positive but statistically insignificant ($p = 0.259$). Due to the short horizon, the Mann–Kendall test did not reach significance, but the direction of the slope remained upward. In the overall period, the increase was again found to be significant ($p < 0.001$). As a result, the forecast part reinforced the actual trend in terms of direction, and the overall line remained upward-sloping, consistent with the green line in Figure 9a.

The radiation variable actual slope was found to be very close to zero and insignificant ($p = 0.727$). It was understood that no significant trend emerged in the long term. Although a small positive slope was predicted in the forecast, it is insignificant ($p = 1.000$). Overall, the line remains horizontal-close and insignificant ($p = 0.107$). Thus, consistent with the visual intuition in Figure 9b, it has been confirmed that radiation behaves stably in the long term.

A significant decrease was observed in the relative humidity variable during the actual period ($p < 0.001$). The slope remained negative in the forecast but was found to be insignificant ($p = 0.133$). Directional consistency was maintained, but it limited the power of the short-horizon Mann–Kendall test. The overall result reconfirmed the significant decrease ($p < 0.001$). This was consistent with the downward slope of the blue/green lines in Figure 9c.

For the Evapotranspiration variable, a positive and significant increase was found in the actual period ($p < 0.001$). Notably, the forecast period was also positive and significant ($p = 0.024$); thus, the increasing trend was statistically supported even in the short term. The overall result also confirmed the significant increase ($p < 0.001$). This finding coincided with the upward blue/green lines observed in Figure 9d and the composite structure sensitive to evapotranspiration's independent variable forecasts.

Due to the inherent nature of climate and hydrological processes, model predictions inevitably contain uncertainty. Therefore, reporting only point values is insufficient to fully reflect the reliability of future projections. Therefore, in this section, the projected evapotranspiration values for the 2025–2030 period are presented based on yearly means, accompanied by an uncertainty analysis. Using a Monte Carlo/resampling approach based on the error distribution derived from past period errors (residuals), prediction intervals (95% confidence intervals) were calculated for each year. It should be made explicit that this procedure resamples only the residuals of the final Model 2 (evapotranspiration) under the fitted LSTM, and does not separately propagate the forecast uncertainty of the first-stage temperature, radiation, and relative-humidity predictions, nor the uncertainty associated with model selection, hyperparameters, the underlying reanalysis data, or non-stationarity. Consequently, the resulting intervals should be regarded as a lower bound on the true predictive uncertainty and are likely narrower than the full uncertainty envelope of future basin evapotranspiration (see Appendix Table A3).

Monte Carlo simulation results show that the upward trend in evapotranspiration levels predicted by the final LSTM-based model for the 2025–2030 period is also maintained on a scenario basis (see Figure 10). A total of 10,000 scenarios were generated for each year using LSTM point estimates and a Monte Carlo/resampling approach based on historical error distribution. The expected (mean) values for the yearly means obtained from these scenarios increased from 3.519 mm/day in 2025 to 3.726 mm/day in 2030. In contrast, the minimum scenario means, interpreted as the expected best-case scenario, ranged from 3.469 to 3.676 mm/day between 2025 and 2030, while the maximum scenarios, representing the expected worst-case scenario, ranged from 3.577 mm/day to 3.784 mm/day. Therefore, although the uncertainty band is relatively narrow (approximately 0.10–0.12 mm/day wide each year), all scenario ranges indicate an upward trend in evapotranspiration. This supports a negative climatic/hydrological picture in terms of increased evaporation and irrigation water demand at the basin scale. (Summary statistics and visual inspection results regarding the distribution of error terms are presented in Appendix Figure A5).

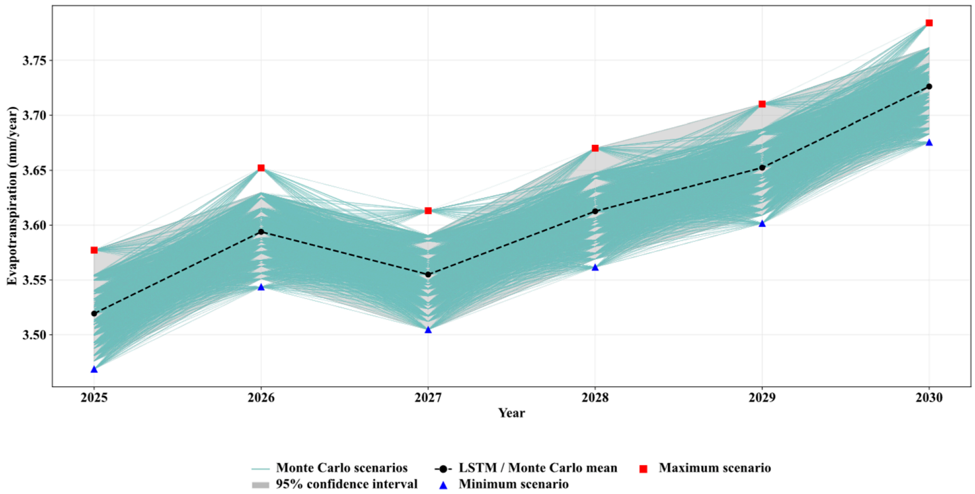


**Figure 10.** Yearly evapotranspiration estimates obtained from LSTM-based Monte Carlo simulations for the 2025–2030 period: range graph for 10,000 scenarios, 95% confidence interval (0.025–0.975 percentiles), LSTM/Monte Carlo mean prediction curve, and expected best (minimum scenario) and worst (maximum scenario) evapotranspiration levels for each year.

## 4. Discussion

The LSTM-based final evapotranspiration model developed in this study demonstrated high explanatory power and low error levels during the validation phase, with values of $R^2$ = 0.941, NSE = 0.991, KGE = 0.974, RMSE = 0.191 mm/day, MAE = 0.131, and MAPE = 4.061 (Table 3). Figure 11 shows these results, labeled "This study (Türkiye, LSTM)," side by side on a world map alongside 15 studies reported in the literature from different countries. For each country, only the model with the best performance in the study and the key performance metrics reported for that model (e.g., $R^2$, NSE, KGE, RMSE, MAE) are presented. Considering that most studies worked with monthly or daily reference/actual evapotranspiration, different climate zones, and different input sets, it should be emphasized that these comparisons should be interpreted in terms of methodology and order of magnitude rather than absolute ranking. This figure is intended only as a qualitative context map: because the cited studies differ in target variable, data source, climate regime, temporal resolution, input set, and validation strategy, the side-by-side metrics must not be read as evidence that the present model outperforms or is directly equivalent to models developed elsewhere. Moreover, the forecasted $ET_0$ is only one term in the closed-basin water balance; the observed decline of Lake Burdur depends additionally on precipitation, surface inflow, groundwater exchange, open-water evaporation, irrigation abstraction, and lake-area feedbacks, so rising $ET_0$ should be read as an indicator of increasing atmospheric water demand rather than a direct quantification of lake water loss.

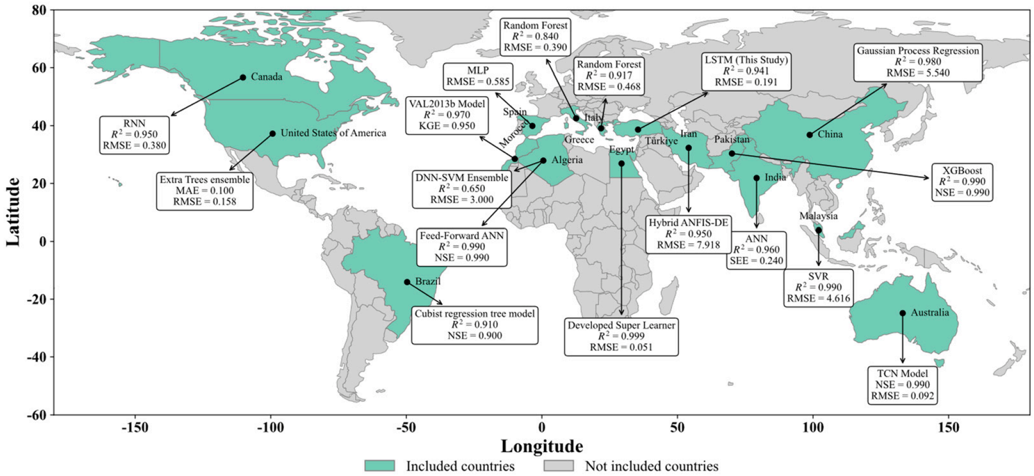


**Figure 11.** Geographic distribution of selected studies on predicting evapotranspiration worldwide using machine learning/deep learning models and performance metrics of the best-performing models (Note: The LSTM model in Türkiye is highlighted as "This study").

From the perspective of deep learning-based approaches, Afzaal et al. (2020) report $R^2$ = 0.950 and RMSE = 0.380 for Canada, while Sarkar et al. (2025) using the TCN architecture in Australia present one of the most successful deep learning examples in the literature with NSE = 0.990 and RMSE = 0.092 [28,29]. In contrast, Achite et al. (2025), who applied a DNN-SVM ensemble for evaporation in Algeria, obtained only $R^2$ = 0.650 and RMSE = 3.000 [30]. The LSTM model developed for Türkiye in this study achieves an accuracy level close to that of Afzaal et al. (2020) and Sarkar et al. (2025), with $R^2$ = 0.941, NSE = 0.991, and RMSE = 0.191, as shown in Figure 11, while producing significantly lower errors than the study by Achite et al. (2025) [28–30]. Therefore, it can be said that the LSTM architecture holds a competitive position within the deep learning family for daily-scale evapotranspiration series with strong seasonality (See Figure 11). At the same time, these high scores should be read with caution: because daily $ET_0$ is strongly seasonal and highly correlated with radiation and temperature, much of the explained variance reflects reproduction of the annual cycle rather than prediction of anomalies or extremes, so performance on de-seasonalized anomalies and extreme high-$ET_0$ episodes is expected to be lower. The study also connects to recent work on machine-learning ET estimation under limited inputs in Mediterranean and data-scarce settings; for example, Bellido-Jimenez et al. (2021) improved reference-ET estimates using intra-daily temperature-based variables in semi-arid Spain, and Stefanidis et al. (2025) compared machine-learning algorithms for potential-ET estimation with limited data at a high-altitude Mediterranean forest [31,32] (see Figure 11).

Comparisons with tree-based and other machine learning models also support a similar picture. For Wang et al. (2022) using XGBoost in Pakistan, Figure 11 shows $R^2$ = 0.990 and NSE = 0.990 values, reporting a performance level nearly identical to this study, particularly in terms of NSE (NSE = 0.991) [33]. Nasir et al. (2025) using the SVR model in Malaysia obtained $R^2$ = 0.990 and RMSE = 4.616, while Elbeltagi et al. (2025) using Gaussian Process Regression (GPR) in China obtained $R^2$ = 0.980 and RMSE = 5.540 [34,35]. Although both studies produced quite high $R^2$ values, their RMSE values were higher than the RMSE = 0.191 value in this study (scale and unit differences should be taken into account) [34,35]. Dias et al. (2021) in Brazil using a Cubist regression tree reported $R^2$ = 0.910 and NSE = 0.900 in Figure 11, while Pagano et al. (2023) in Italy modeling daily actual evapotranspiration using Random Forest reported $R^2$ = 0.840 and RMSE = 0.390 [36,37]. Patel et al. (2025) in the US, using the Extra Trees ensemble, achieved a low error level with MAE = 0.100 and RMSE = 0.158 [38]. These values show that the LSTM model in Türkiye, with metrics of $R^2$ = 0.941, NSE = 0.991, RMSE = 0.191, and MAE = 0.131 (Figure 11), offers the same level of accuracy as tree-based models and, when compared to some studies, a more consistent performance, particularly in terms of NSE and KGE.

Similar results were observed when compared with artificial neural networks, ANFIS, and hybrid/ensemble approaches. In the study by Abraham and Mohan (2023) using PCA-enhanced ANN in India, $R^2$ = 0.960 and SEE = 0.240 were reported in Figure 11, while Aghelpour et al. (2022) reported $R^2$ = 0.950 and RMSE = 7.918 [39,40]. Achite et al. (2022), who studied reference evapotranspiration in Algeria using feed-forward ANN and GEP models, obtained $R^2$ = 0.990 and NSE = 0.990, while Acharki et al. (2025), who used the VAL2013b hybrid model in Morocco, achieved $R^2$ = 0.970 and KGE = 0.950 [41,42]. Aly et al. (2024), who developed the "developed super learner" ensemble model in Egypt, reported one of the highest accuracy levels in the literature, with $R^2$ = 0.999 and RMSE = 0.051, as shown in Figure 11 [43]. When evaluated together with these studies, the KGE = 0.974 and NSE = 0.991 values of the LSTM model in this study (Figure 11) are at the same level as the highly hydrologically consistent models such as Acharki et al. (2025) [42], Achite et al. (2022) [41], and Wang et al. (2022) [33], which are the models with high hydrological consistency. In terms of error magnitude, it is relatively close to powerful ensemble and deep learning frameworks such as Aly et al. (2024) [43] and Sarkar et al. (2025) [29].

## 5. Conclusions

In this study, a performance-based, nested prediction framework combining statistical, machine learning, and deep learning methods was developed and evaluated to estimate daily evapotranspiration values for Lake Burdur in the near future (2025–2030). Within Model 1, SARIMAX, XGBoost, and LSTM algorithms were systematically compared for temperature, radiation, and relative humidity series using 37-layer rolling-origin time series cross-validation; at the validation stage, LSTM ($R^2$ = 0.967), XGBoost ($R^2$ = 0.845), and SARIMAX ($R^2$ = 0.812) were determined to be the most successful algorithms for temperature, radiation, and relative humidity, respectively. In Model 2, estimation was performed using the best projections of these variables as independent variables, and the LSTM architecture achieved $R^2$ = 0.941, NSE = 0.991, KGE = 0.974, RMSE = 0.191 mm/day, MAE = 0.131, and MAPE = 4.061, demonstrating a significant superiority over other algorithms. Trend analyses based on yearly means indicate that the significant increase in temperature and evapotranspiration observed in the historical period, along with the marked decrease in relative humidity, will continue in the 2025–2030 period, while there may not be a significant break in the radiation variable at the general level.

The proposed nested modeling framework ensures that model selection is appropriate for the structure of the variable (statistical time series models for relatively linear and noisy processes, tree-based ML for complex but moderately noisy processes, LSTM for series with strong seasonality and high autocorrelation) can significantly improve prediction accuracy compared to single-stage model approaches. Monte Carlo uncertainty analysis based on past error distributions revealed that 10,000 different scenarios point to an increasing trend in evapotranspiration between 2025 and 2030, supporting a negative climatic/hydrological frame in terms of increased evaporation and irrigation water demand at the basin scale. Future studies should focus on integrating in situ station observations with satellite observations; developing hybrid process-data-driven models that explicitly represent land use dynamics and irrigation feedback; and evaluating simpler hyperparameter optimization strategies that improve the accuracy-computational cost trade-off. Furthermore, the transferability of the proposed nested framework should be tested in other Mediterranean and semi-arid basins to assess its generalizability under different climatic and physiographic conditions.

## 6. Limitations

This study is based on meteorological explanatory variables derived from a single reanalysis grid point representing the Burdur Lake basin. Although reanalysis products provide spatially and temporally consistent coverage, a single grid point may not fully capture lake–land microclimate and local heterogeneity (e.g., coastal gradient, topography-modulated processes, and moisture/radiation variability). Therefore, part of the reported performance may reflect the internal consistency of the reanalysis dataset rather than the full variation in actual field observations. Furthermore, due to the limited availability/accessibility of long-term, quality-controlled in situ station observations during the study period, a systematic station-based comparison for reanalysis inputs and the derived ET0 series could not be performed. This limits the assessment of absolute bias and restricts the generalizability of the selected grid representation to situations where it may deviate from local conditions. Additionally, this study employs a nested forecasting design, where the intermediate forecasts generated by Model 1 are fed into Model 2. This approach carries the risk of propagating errors originating from Model 1 into Model 2, particularly during the 2025–2030 forecast period, and potentially amplifying them under certain conditions. Therefore, the reported performance and forecast uncertainty are influenced not only by the structure of Model 2, but by the entire two-stage chain (for information flow and potential error propagation schematically, see Appendix Table A3). Finally, the 2025–2030 forecast horizon is relatively short compared to climate change timescales; therefore, the results should be interpreted as near-term and scenario-conditional forecasts rather than long-term climate projections, bearing in mind that uncertainty increases rapidly as the horizon extends.

**Author Contributions:** Conceptualization, M.G., N.O.Ü., D.Y. (Doğan Yıldız) and D.Y. (Dursun Yıldız); methodology, M.G. and N.O.Ü.; software, M.G. and N.O.Ü.; validation, M.G., N.O.Ü., D.Y. (Doğan Yıldız) and D.Y. (Dursun Yıldız); formal analysis, M.G. and N.O.Ü.; investigation, M.G. and N.O.Ü.; resources, M.G. and N.O.Ü.; data curation, M.G. and N.O.Ü.; writing—original draft preparation, M.G. and N.O.Ü.; writing—review and editing, D.Y. (Doğan Yıldız) and D.Y. (Dursun Yıldız); visualization, M.G. and N.O.Ü.; supervision, D.Y. (Doğan Yıldız) and D.Y. (Dursun Yıldız); project administration, M.G. and N.O.Ü. The interpretation of the analytical findings and their evaluation in terms of water policies were supervised and reviewed by D.Y. (Doğan Yıldız) and D.Y. (Dursun Yıldız). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The meteorological and evapotranspiration data used in this study were obtained from the publicly available Open-Meteo platform. The processed datasets and analysis outputs generated during the current study are available from the corresponding author upon reasonable request [44].

**Acknowledgments:** The authors would like to thank the Open-Meteo platform for providing open-access meteorological and evapotranspiration data used in this study.

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

## Appendix A

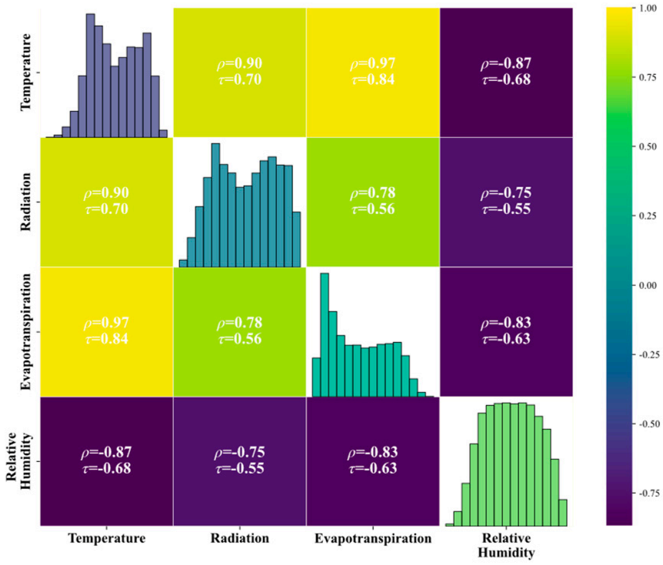


**Figure A1.** Distribution histograms, Spearman's rho ($\rho$) and Spearman's Kendall ($\tau$) correlation coefficients of climatic variables.

**Table A1.** ADF, KPSS (Trend), and KPSS (Level) Test Statistics and Stationarity Decisions for Key Meteorological Variables (Note: Reported $p$-values for ADF, KPSS (Trend), and KPSS (Level) are evaluated at $\alpha = 0.05$; decisions are made at this significance level (ADF $H_0$: unit root; KPSS $H_0$: stationarity)).

| Variable | ADF [1] | | KPSS (Level) [2] | | KPSS (Trend) [3] | |
|---|---|---|---|---|---|---|
| | Test | $p$-Value | Test | $p$-Value | Test | $p$-Value |
| Evapotranspiration | −10.960 | 0.000 | 0.035 | 0.100 | 0.004 | 0.100 |
| Temperature | −10.291 | 0.000 | 0.140 | 0.100 | 0.004 | 0.100 |
| Radiation | −10.225 | 0.000 | 0.006 | 0.100 | 0.005 | 0.100 |
| Relative Humidity | −9.039 | 0.000 | 0.254 | 0.100 | 0.010 | 0.100 |

[1] $ADF_{Critical}$ is −2.862. [2] KPSS $(Level)_{Critical}$ is 0.463. [3] KPSS $(Trend)_{Critical}$ is 0.146.

**Table A2.** Residual diagnosis tests for selected SARIMA(X) orders for Model 1 and Model 2 (Ljung-Box: multiple lags 10-20-30-60-365, KPSS: level/trend, White & Breusch–Pagan): $p$-value summary.

| Predicted Variable | Best Fit | | Model Generation Residual Tests $p$-Value Statistics [1] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | AR | MA | Ljung-Box $Q$ (LAG) | | | | | KPSS | | LM | |
| | $(p, P)$ | $(q, Q)$ | 10 | 20 | 30 | 60 | 365 | Level | Trend | WH [2] | BP [3] |
| Temperature | (2, 1) | (2, 1) | 0.312 | 0.284 | 0.351 | 0.408 | 0.529 | 0.453 | 0.612 | 0.327 | 0.411 |
| Radiation | (2, 0) | (2, 1) | 0.071 | 0.093 | 0.128 | 0.214 | 0.447 | 0.158 | 0.232 | 0.067 | 0.082 |

**Table A2.** *Cont.*

| Predicted Variable | Best Fit | | Model Generation Residual Tests $p$-Value Statistics [1] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | AR | MA | Ljung-Box $Q$ (LAG) | | | | | KPSS | | LM | |
| | ($p$, $P$) | ($q$, $Q$) | 10 | 20 | 30 | 60 | 365 | Level | Trend | WH [2] | BP [3] |
| Relative Humidity | (2, 1) | (1, 0) | 0.229 | 0.318 | 0.364 | 0.392 | 0.541 | 0.379 | 0.544 | 0.258 | 0.303 |
| Evapotranspiration | (2, 1) | (2, 1) | 0.058 | 0.077 | 0.141 | 0.238 | 0.486 | 0.421 | 0.695 | 0.122 | 0.176 |

[1] Statistical significance evaluated at $\alpha = 0.05$. [2] WH: White Noise & Heteroscedasticity Test. [3] BP: Breusch–Pagan Test.

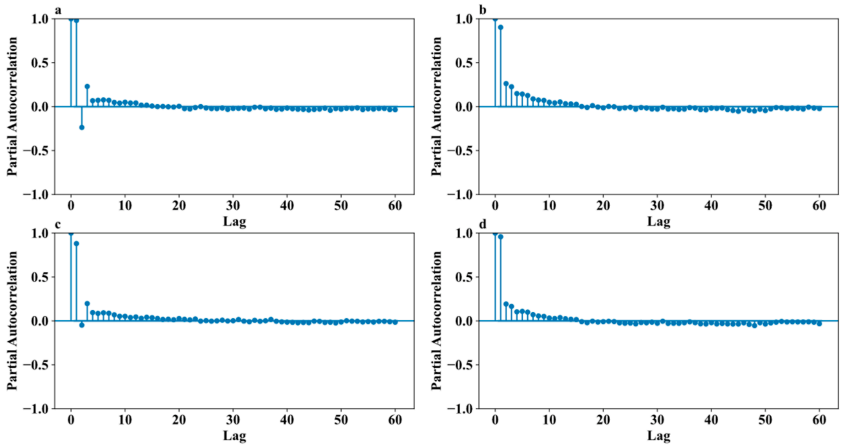


**Figure A2.** Partial autocorrelation (PACF) plots for independent (Model 1) and dependent (Model 2) variables: (**a**) temperature (Model 1), (**b**) radiation (Model 1), (**c**) relative humidity (Model 1), (**d**) evapotranspiration (Model 2).

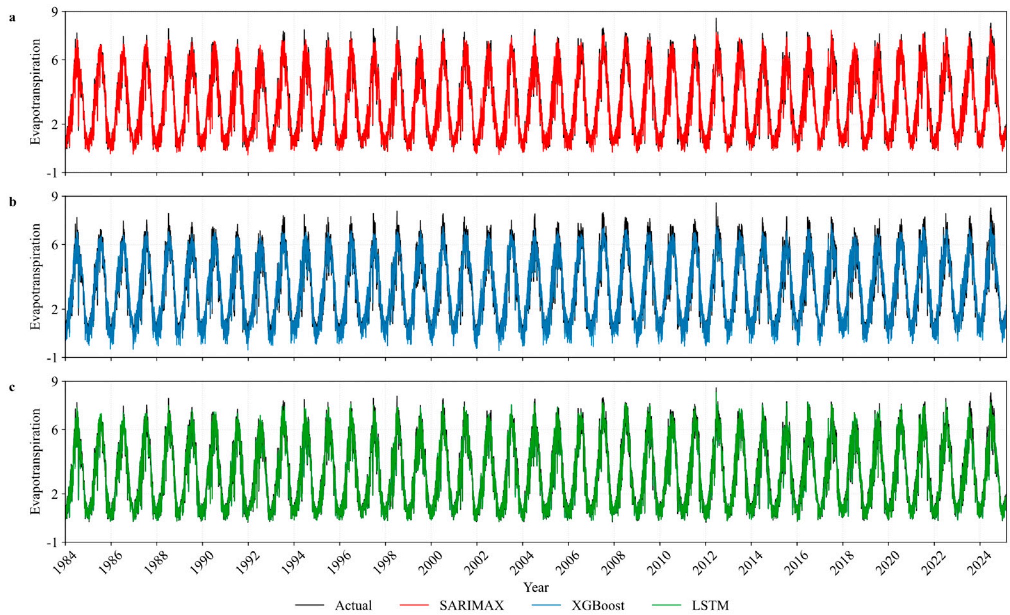


**Figure A3.** Panel (**a**) shows the observed values of daily evapotranspiration (black) and the predictions generated by the SARIMA(X) model for the period 1984–2024. Panel (**b**) shows the observations and XGBoost predictions for the same period. Panel (**c**) shows the observations and LSTM predictions. The vertical axis represents years, and the horizontal axis represents daily evapotranspiration values; in each panel, black lines represent the actual series, while colored lines represent the prediction series of the corresponding model, visually demonstrating the extent to which the models capture both level and seasonal fluctuations. Particularly in panel c, it can be observed that the green prediction curves of the LSTM model largely cover the black observation curves and show an almost overlapping fit with the actual series.

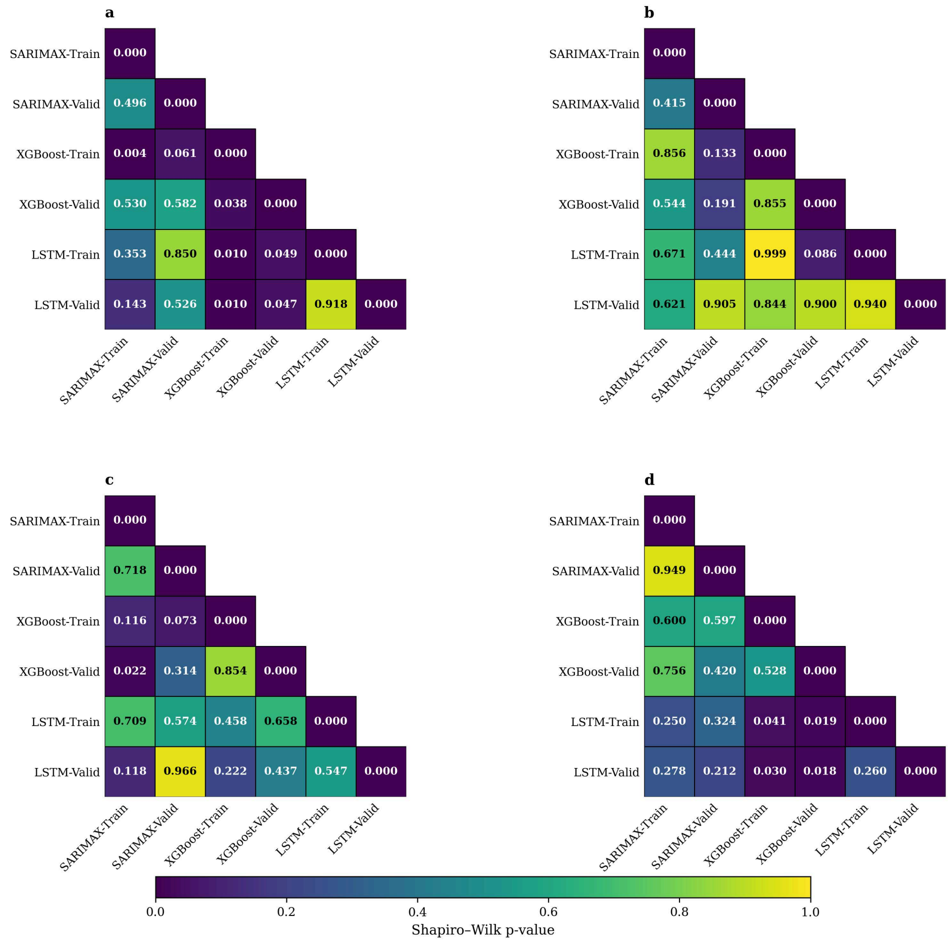


**Figure A4.** Shapiro–Wilk *p*-value matrices for the normality assessment of paired $R^2$ differences across 37 rolling-origin folds for temperature (**a**), radiation (**b**), relative humidity (**c**), and evapotranspiration (**d**).

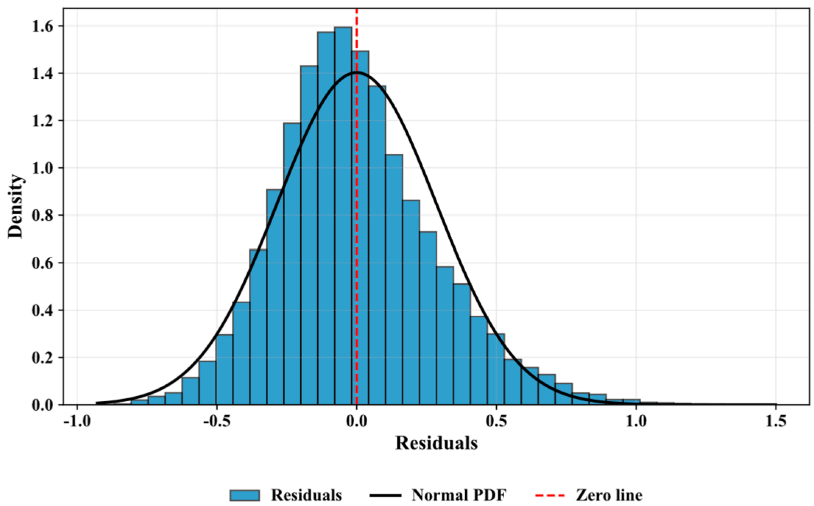


**Figure A5.** Summary statistics obtained for 14,976 residuals (mean = 0.000, standard deviation = 0.284, minimum = −0.93, maximum = 1.50) and the Shapiro–Wilk normality test result ($W_{Test}$ = 0.9836, $p = 4.87 \times 10^{-38}$) indicates that the assumption of a strictly normal distribution is rejected ($p < 0.01$). However, the approximately symmetric structure of the distribution and its visual approximation to the theoretical normal curve make it practically reasonable to use a normal distribution-based approach for error terms in the Monte Carlo uncertainty analysis.

**Table A3.** During the forecast period, Model 2 does not take the "observed" intermediate values; instead, it uses the estimates produced by Model 1 as input. Therefore, the uncertainty associated with Model 1 is reflected as an additional component in the input uncertainty of Model 2.

| Stage | Model | Inputs | Output | Purpose | Key Risk |
|---|---|---|---|---|---|
| 1 | Model 1 | Reanalysis-based meteorological inputs (sliding window) | Interim estimate (Output to be transferred to Model 2) | (Output to be transferred to 2) Generate intermediate input to be used in Model 2 during the forecast period | The error in the interim estimate can be carried forward to the next stage. |
| 2 | Model 2 | Interim estimates of Model 1 (and any additional auxiliary variables) | $\hat{ET}_0$ | Generating the final $\hat{ET}_0$ forecast | Total error may increase because errors in Model 1 propagate to the input (error propagation). |

## References


1. IPCC. *Climate Change 2021: The Physical Science Basis*; Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Cambridge University Press: Cambridge, UK, 2021. [CrossRef]
2. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements (FAO Irrigation and Drainage Paper No. 56)*; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998.
3. Rotenberg, E.; Tatarinov, F.; Muller, J.D.; Yakir, D. Evapotranspiration saturation amplifies climate sensitivity of terrestrial water yield. *Nat. Commun.* **2025**, *14*, 6657. [CrossRef] [PubMed]
4. 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]
5. Shiri, J. Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran. *Agric. Water Manag.* **2017**, *188*, 101–114. [CrossRef]
6. Ayaz, A.; Rehana, S.; Rajesh, M.; Singh, S.K. Estimation of reference evapotranspiration using machine learning models with limited data. *AIMS Geosci.* **2021**, *7*, 268–290. [CrossRef]
7. Raza, A.; Shoaib, M.; Faiz, M.A.; Baig, F.; Khan, M.M.; Ullah, M.K.; Zubair, M. Comparative assessment of reference evapotranspiration estimation using conventional method and machine learning algorithms in four climatic regions. *Pure Appl. Geophys.* **2020**, *177*, 4479–4508. [CrossRef]
8. Tikhamarine, Y.; Malik, A.; Pandey, K.; Sammen, S.S.; Gamane, D.S.; Heddam, S.; Kisi, O. Monthly evapotranspiration estimation using optimal climatic parameters: Efficacy of hybrid support vector regression integrated with whale optimization algorithm. *Environ. Monit. Assess.* **2020**, *192*, 696. [CrossRef] [PubMed]
9. Carter, C.; Liang, S. Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 86–92. [CrossRef]
10. 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]
11. Costa, T.S.; Filgueiras, R.; Dos Santos, R.A.; Cunha, F. Actual evapotranspiration by machine learning and remote sensing without the thermal spectrum. *PLoS ONE* **2023**, *18*, e0285535. [CrossRef] [PubMed]
12. Du, C.; Jiang, S.; Chen, C.; Guo, Q.; He, Q.; Zhan, C. Machine learning-based estimation of daily cropland evapotranspiration in diverse climate zones. *Remote Sens.* **2024**, *16*, 730. [CrossRef]
13. Melton, F.S.; Huntington, J.; Grimm, R.; Herring, J.; Hall, M.; Rollison, D.; Erickson, T.; Allen, R.; Anderson, M.; Fisher, J.B.; et al. OpenET: Filling a critical data gap in water management for the western United States. *J. Am. Water Resour. Assoc.* **2022**, *58*, 971–994. [CrossRef]
14. Chukalla, A.D.; Mul, M.L.; van der Zaag, P.; van Halsema, G.; Mubaya, E.; Muchanga, E.; den Besten, N.; Karimi, P. A framework for irrigation performance assessment using WaPOR data: The case of a sugarcane estate in Mozambique. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 2759–2778. [CrossRef]
15. Amini, A.; Emami, S.; Dehghanisanij, H. Participatory evaluation of an irrigation decision support system for water-saving and productivity gains in Lake Urmia Basin. *Sci. Rep.* **2025**, *15*, 42480. [CrossRef] [PubMed]
16. Cahn, M.; Johnson, L.; Benzen, S. Evapotranspiration-based irrigation management effects on yield and water productivity of summer cauliflower on the California Central Coast. *Horticulturae* **2025**, *11*, 322. [CrossRef]

17. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [CrossRef] [PubMed]
18. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. [CrossRef]
19. Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431. [CrossRef]
20. Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom.* **1992**, *54*, 159–178. [CrossRef]
21. Shah, V.; Patel, N.; Shah, D.; Swain, D.; Mohanty, M.; Acharya, B.; Gerogiannis, V.C.; Kanavos, A. Forecasting maximum temperature trends with SARIMAX: A case study from Ahmedabad, India. *Sustainability* **2024**, *16*, 7183. [CrossRef]
22. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018. Available online: https://otexts.com/fpp2/ (accessed on 23 June 2026).
23. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
24. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization (AdamW). *arXiv* **2019**. [CrossRef]
25. Yao, Y.; Rosasco, L.; Caponnetto, A. On early stopping in gradient descent learning. *Constr. Approx.* **2007**, *26*, 289–315. [CrossRef]
26. Kratzert, F.; Klotz, D.; Shalev, G.; Klambauer, G.; Hochreiter, S.; Nearing, G. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 5089–5110. [CrossRef]
27. Lees, T.; Reece, S.; Kratzert, F.; Klotz, D.; Gauch, M.; De Bruijn, J.; Sahu, R.K.; Greve, P.; Slater, L.; Dadson, S.J. Hydrological concept formation inside long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 3079–3101. [CrossRef]
28. Afzaal, H.; Farooque, A.A.; Abbas, F.; Acharya, B.; Esau, T. Computation of Evapotranspiration with Artificial Intelligence for Precision Water Resource Management. *Appl. Sci.* **2020**, *10*, 1621. [CrossRef]
29. Sarkar, S.S.; Bedi, J.; Jain, S. A deep learning-based framework for enhanced reference evapotranspiration estimation: Evaluating accuracy and forecasting strategies. *Sci. Rep.* **2025**, *15*, 15136. [CrossRef] [PubMed]
30. Achite, M.; Kumar, M.; Elshaboury, N.; Srivastava, A.; Elbeltagi, A.; Salem, A. Comparative assessment of standalone and hybrid deep neural networks for modeling daily pan evaporation in a semi-arid environment. *Sci. Rep.* **2025**, *15*, 20179. [CrossRef] [PubMed]
31. Bellido-Jiménez, J.A.; Estévez, J.; García-Marín, 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]
32. Stefanidis, S.; Ioannou, K.; Proutsos, N.; Karmiris, I.; Stefanidis, P. Comparative analysis of machine learning algorithms for potential evapotranspiration estimation using limited data at a high-altitude Mediterranean forest. *Atmosphere* **2025**, *16*, 851. [CrossRef]
33. Wang, J.; Raza, A.; Hu, Y.; Buttar, N.A.; Shoaib, M.; Saber, K.; Li, P.; Elbeltagi, A.; Ray, R.L. Development of Monthly Reference Evapotranspiration Machine Learning Models and Mapping of Pakistan-A Comparative Study. *Water* **2022**, *14*, 1666. [CrossRef]
34. Nasir, M.A.M.; Harun, S.; Zainuddin, Z.M.; Kamal, R.; Rose, F.Z.C. Machine learning techniques for reference evapotranspiration and rice irrigation requirements prediction: A case study of Kerian irrigation scheme, Malaysia. *Paddy Water Environ.* **2025**, *23*, 671–683. [CrossRef]
35. Elbeltagi, A.; Srivastava, A.; Cao, X.; Gautam, V.K.; Zerouali, B.; Aslam, M.R.; Salem, A.; Emami, H.; Elsadek, E.A. Bayesian-optimized machine learning boosts actual evapotranspiration prediction in water-stressed agricultural regions of China. *Sci. Rep.* **2025**, *15*, 37548. [CrossRef] [PubMed]
36. Dias, S.H.B.; Filgueiras, R.; Filho, E.I.F.; Arcanjo, G.S.; da Silva, G.H.; Mantovani, E.C.; da Cunha, F.F. Reference evapotranspiration of Brazil modeled with machine learning techniques and remote sensing. *PLoS ONE* **2021**, *16*, e0245834. [CrossRef] [PubMed]
37. Pagano, A.; Amato, F.; Ippolito, M.; De Caro, D.; Croce, D.; Motisi, A.; Provenzano, G.; Tinnirello, I. Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation. *Ecol. Inform.* **2023**, *76*, 102133. [CrossRef]
38. Patel, A.; Ali, S.T.; Pandey, M.K. Estimation of reference evapotranspiration using ensemble machine learning models based on regional scenario. *Appl. Water Sci.* **2025**, *15*, 307. [CrossRef]
39. Abraham, M.; Mohan, S. ANN-based PCA to predict evapotranspiration: A case study in India. *AQUA-Water Infrastruct. Ecosyst. Soc.* **2023**, *72*, 1145–1163. [CrossRef]
40. Aghelpour, P.; Varshavian, V.; Khodamorad Pour, M.; Hamedi, Z. Comparing three types of data-driven models for monthly evapotranspiration prediction under heterogeneous climatic conditions. *Sci. Rep.* **2022**, *12*, 17363. [CrossRef] [PubMed]
41. Achite, M.; Jehanzaib, M.; Sattari, M.T.; Toubal, A.K.; Elshaboury, N.; Wałęga, A.; Krakauer, N.; Yoo, J.-Y.; Kim, T.-W. Modern Techniques to Modeling Reference Evapotranspiration in a Semiarid Area Based on ANN and GEP Models. *Water* **2022**, *14*, 1210. [CrossRef]

42. Acharki, S.; Raza, A.; Vishwakarma, D.K.; Amharref, M.; Bernoussi, A.S.; Singh, S.K.; Al-Ansari, N.; Dewidar, A.Z.; Al-Othman, A.A.; Mattar, M.A. Comparative assessment of empirical and hybrid machine learning models for estimating daily reference evapotranspiration in sub-humid and semi-arid climates. *Sci. Rep.* **2025**, *15*, 2542. [CrossRef] [PubMed]
43. Aly, M.S.; Darwish, S.M.; Aly, A.A. High performance machine learning approach for reference evapotranspiration estimation. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 689–713. [CrossRef]
44. Open-Meteo. Historical Weather API [Computer Software/API]. 2025. Available online: https://open-meteo.com/en/docs/historical-weather-api (accessed on 3 April 2025).

**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.