Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated Multi-Horizon Forecasting of Hourly Reference Evapotranspiration using Statistical, Tabular, Ensemble, and Foundation Models

Sowmya M R

sowmya@ssc.edu.in

Sree Sankara College https://orcid.org/0000-0001-7866-8164

Santosh Kumar M B

Cochin University of Science and Technology

Research Article

Keywords: Reference evapotranspiration, AutoML, AutoGluon-Timeseries, Short-term multi horizon hourly forecasting, Chronos2

Posted Date: June 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9792214/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Automated Multi-Horizon Forecasting of Hourly Reference Evapotranspiration using Statistical, Tabular, Ensemble, and Foundation Models

Sowmya M R[1] and Santosh Kumar M B[2]

[1] Department of Computer Science, Sree Sankara College, Kalady, Ernakulam. Kerala, 683574, India
ORCID ID: 0000-0001-7866-8164

[2] Division of Information Technology, School of Engineering, Cochin University of Science and Technology, Kochi, Kerala, 682022, India
ORCID ID: 0000-0002-1893-8414

Corresponding Author: Sowmya M R, Department of Computer Science, Sree Sankara College, Kalady, Ernakulam, Kerala, 683574, India
E-mail: sowmya@ssc.edu.in, sowmyamr.ssc@gmail.com

Abstract

Hourly reference evapotranspiration (ET_0) forecasting plays a critical role in irrigation scheduling, water resource planning, and climate-resilient agricultural management. Short-term multi-horizon forecasting of hourly ET_0 remains relatively underexplored, particularly using automated forecasting frameworks and emerging time-series foundation models. This study proposes an automated multi-horizon forecasting framework for hourly ET_0 prediction using statistical, tabular, ensemble, deep-learning, and foundation forecasting paradigms implemented through the AutoGluon TimeSeries (AG-TS) framework. Hourly ET_0 values were computed using the ASCE Penman–Monteith equation from meteorological observations collected between October 2015 and December 2025. Forecasting experiments were conducted for four operational horizons (1-h, 6-h, 24-h, and 168-h) using SeasonalNaive, Theta, ETS, RecursiveTabular, DirectTabular, WeightedEnsemble, Temporal Fusion Transformer (TFT), and Chronos2 models. Results revealed horizon-dependent model behaviour, with Chronos2 achieving superior performance at the 1-h and 168-h forecasting horizons, while WeightedEnsemble demonstrated better predictive capability at the 6-h and 24-h horizons. Comparative analysis further indicated that Chronos2 maintained competitive forecasting accuracy with comparatively moderate computational demand, whereas TFT incurred substantially higher training costs despite rapid inference capability. These findings demonstrate the potential of automated forecasting frameworks for robust multi-horizon ET_0 prediction and highlight the emerging applicability of pretrained time-series foundation models for operational hydrological and agro-environmental forecasting tasks.

Keywords: Reference evapotranspiration, AutoML, AutoGluon-Timeseries, Short-term multi-horizon hourly forecasting, Chronos2

1. Introduction

Reference evapotranspiration (ET_0) is an important meteorological component that contributes to effective agricultural water management and irrigation, hydrological modelling, and climate-resilient crop production systems (Goyal et al. 2023). The extreme climate variability and water scarcity demand an accurate estimation and forecasting of ET_0 as it optimizes irrigation water use, improves agricultural productivity, and supports sustainable water-resource management (Chen et al. 2025). Recently, the application of precision agriculture has contributed to the necessity to have short-term forecast of the ET_0 in order to support operational decisions at short-time intervals (Zhou et al. 2026). This is particularly applicable for regions where water availability is limited and there is an acute need to manage water consumption more effectively.

Traditionally, ET_0 estimation has been performed using empirical equations, statistical models, and process-based hydrological approaches. Conventional methods such as Hargreaves-Samani, Turc and Penman-Monteith equations are widely utilized for estimating ET_0, even in data-scarce regions(Xiang et al. 2020). However, these traditional approaches are often limited by the nonlinear interactions between climates, uncertainty in meteorological inputs, and the dynamic nature of atmospheric processes(Chia et al. 2020). Moreover, these methods are primarily focused on estimating ET_0 from historical and current meteorological observations, and extensive research has already been conducted in this area, making it a relatively saturated field(Goyal et al. 2023). In contrast, ET_0 forecasting, which involves predicting future ET_0

values, has emerged as an important research domain with significant potential for improving real-time irrigation scheduling and supporting better water management decisions(Gao et al. 2025, 2026; Tang et al. 2025; Yan et al. 2026). However, ET_0 forecasting studies remain comparatively limited in the literature when compared to ET_0 estimation studies(Yan et al. 2025; Gupta et al. 2026).

Time-series forecasting has numerous important applications in modern water resources management(Sang 2013; Willard et al. 2025), including environmental protection, water supply planning, water quality management, irrigation system operation, hydropower generation, and sustainable utilization of water resources. Consequently, research attention has gradually shifted toward advanced statistical time-series models, machine learning (ML) models, and deep learning (DL) models for ET_0 forecasting. Some early findings demonstrated the potential of statistical learning approaches such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) for weekly and daily ET_0 forecasting(Afifi 2020; Aghelpour and Norooz-Valashedi 2022), particularly when historical ET_0 observations exhibited strong temporal dependencies. However, these methods generally struggled to capture highly nonlinear and nonstationary temporal dynamics, especially for multi-step forecasting horizons. As ML techniques advances, many researchers attempted to forecast ET_0 using artificial neural networks (Landeras et al. 2009), relevance vector machines in standard and hybrid forms (Torres et al. 2011; Bachour et al. 2016), Gaussian process regression (Karbasi 2018), and ensemble learning approaches (Nourani et al. 2020). In most cases, ML-based forecasting models were proved to be superior to conventional statistical models in terms of accuracy and effectiveness, especially under complex climatic conditions and limited-input scenarios. Hybrid approaches combining decomposition techniques such as wavelet transform (Karbasi 2018), variational mode decomposition (VMD) (Malik et al. 2022; Lu et al. 2023; Lee et al. 2024; Chen et al. 2025), empirical mode decomposition (Karbasi et al. 2023), and feature-selection strategies further improved forecasting robustness by reducing non-stationarity and extracting hidden temporal patterns from ET_0 series (Ahmed et al. 2022; Karbasi et al. 2023, 2024). In particular, decomposition-hybrid frameworks such as Wavelet-GPR (Karbasi 2018), VMD-BPNN (Lu et al. 2023), and MVMD-RR-KELM (Malik et al. 2022) achieved promising forecasting performance for short- and medium-term ET_0 prediction.

The more recent DL techniques have dominated ET_0 forecasting investigations owing to their ability to capture complex nonlinear temporal relationships. Time lagged-recurrent neural networks (Proias et al. 2020), long short-term memory (LSTM) networks (Alibabaei et al. 2020), gated recurrent units(GRU) (Zhang et al. 2025), bidirectional LSTM (Qian et al. 2024; Yan et al. 2025), convolutional neural networks (CNNs)(de Oliveira e Lucas et al. 2020; Ferreira and da Cunha 2020a), and ConvLSTM frameworks (Babaeian et al. 2022) have been extensively examined for multi-step ET_0 forecasting across various climate zones. Several studies reported improved forecasting accuracy using hybrid CNN-LSTM (Wang et al. 2023; Li et al. 2024)and attention-based architectures(Tang et al. 2025), while others demonstrated the effectiveness of transfer learning and spatiotemporal forecasting strategies (Bedi 2022; Ye et al. 2025). Recent research also introduced transformer-based (Yan et al. 2026) and attention-enhanced (Gao et al. 2025, 2026) forecasting systems for ET_0 and evapotranspiration prediction, indicating a gradual transition toward more advanced sequence-learning architectures. Meanwhile, explainable artificial intelligence techniques such as SHAP analysis

have started emerging to improve model interpretability and identify dominant climatic drivers affecting ET_0 dynamics (Tang et al. 2025).

Nevertheless, there still exist several crucial deficiencies in the recent literature of forecasting ET_0. First, the majority of existing studies primarily focus on daily, weekly, or monthly forecasting scales, whereas short-term hourly ET_0 forecasting remains comparatively underexplored. Hourly forecasting is operationally important for real-time irrigation scheduling, intraday water-allocation planning, and adaptive smart-irrigation systems, particularly under rapidly changing atmospheric conditions. Second, most existing forecasting frameworks rely heavily on handcrafted architectures, climate-specific tuning, decomposition preprocessing, and extensive feature engineering. Although these approaches may improve forecasting accuracy, they often reduce scalability, increase computational complexity, and limit generalization across locations and climatic conditions. Finally, many recent studies utilize multiple weather factors as input variables that may not always be readily available in operational agricultural environments, especially in data-scarce regions.

Concurrently, recent developments in automated machine learning (AutoML) and pretrained time-series foundation models have introduced new opportunities for scalable and generalized forecasting systems. AutoML frameworks, such as Auto sklearn, H2O AutoML, Auto keras, TPOT, and AutoGluon reduce the need for manual architecture engineering, hyperparameter tuning, and model-selection procedures, thereby enabling efficient deployment of forecasting pipelines(Mangalath Ravindran et al. 2022). In parallel, pretrained time-series foundation models such as Chronos, Chronos-2, TimesFM, and TimeGPT have recently demonstrated strong forecasting performance across multiple domains including finance(Asher Marconi 2025), energy systems(Park et al. 2025), environmental monitoring, and healthcare(Zhang et al. 2026). These models leverage large-scale pretraining to learn generalized temporal representations capable of supporting zero-shot and few-shot forecasting tasks(Ansari et al. 2024). Although pretrained time-series foundation models have recently shown promising performance across several forecasting domains, their application to operational short-term hourly ET_0 forecasting remains absent in the current literature.

To the best of our knowledge, pretrained time-series foundation models have not yet been systematically investigated for univariate short-term hourly ET_0 forecasting within an operational agro-hydrological context. Furthermore, comparative benchmarking between foundation, ensemble, tabular, deep-learning and conventional statistical forecasting paradigms across multiple operational forecasting horizons remains largely unexplored. Addressing these gaps is important for evaluating whether pretrained temporal representations can enhance forecasting robustness while minimizing reliance on handcrafted architectures and extensive meteorological inputs.

Hence the present study aims to

1. Develop a multi-horizon hourly ET_0 forecasting framework using the AutoML architecture, AutoGluon.
2. Compare the forecasting performance of foundation, ensemble, tabular, deep learning and statistical forecasting models across 1 hour (1-h), 6 hours (6-h), 24 hours (24-h), and 168 hours (168-h) forecasting horizons.
3. Evaluate the operational applicability and computational efficiency of heterogenous forecasting paradigms for hourly ET_0 prediction.

4. Investigate the potential capability of pretrained time-series foundation models for operational short-term hourly ET_0 forecasting under the investigated dataset conditions.

2. Materials and Methods

2.1 Study area

Advanced Centre for Atmospheric Radar Research (ACARR) is an atmospheric research centre, which was established in 2015 at Cochin University of Science and Technology (CUSAT), Kochi, Kerala, India. ACARR aims to conduct internationally recognised research in atmospheric science in the tropical monsoon region. The station is equipped with advanced ground-based observational instruments, including a state-of-the-art Automatic Weather Station, which provides continuous measurements of key meteorological variables This facilitates the Indian researchers to incorporate the data in to their operational weather prediction techniques and related modelling procedures (Mangalath Ravindran et al. 2022).

ACARR is situated in Kalamassery, Kochi, Kerala, in the southwestern part of India, at 10.04° N latitude and 76.33° E longitude, with an elevation of 33 m above mean sea level. Kochi is a tropical coastal station located on the windward side of the Western Ghats mountain range in Kerala, approximately 10 km inland from the Arabian Sea coast. It experiences a humid and tropical climate with significant rainfall during most of the year with four seasons. The wind tends to be north-easterly/easterly during the pre-monsoon, post-monsoon, and winter seasons and during the monsoon, it is south-westerly. The ET_0 peaks during the pre-monsoon and diminishes throughout the monsoon period. The annual average temperature is 26.5 ^{0}C, the average relative humidity is 74% and the average yearly rainfall is 2882.9 mm (Mangalath Ravindran et al. 2022).

2.2 Data Set and Data Pre-Processing

The hourly meteorological data such as temperature, humidity, net solar radiation, and wind speed from October 2015 to December 2025 were collected from ACARR database. Around 89000 data records were passed through American Society of Civil Engineers - Penman Monteith equation (APME) to estimate hourly ET_0 (Allen et al. 2006), which is used as the input time series data of the present study. APME is a variant of Penman-Monteith equation for estimating ET_0 on an hourly basis (Ferreira and da Cunha 2020b) and is given by (1).

$$ETo = \frac{0.408\,\Delta\,(R_n - G) + \gamma\left(\frac{37}{T_m + 273}\right)u_2(e_s - e_a)}{\Delta + \gamma(1 + C_h u_2)} \quad (1)$$

where ETo is the reference evapotranspiration (mm/hour), Δ is the slope of vapour pressure curve (kPa/^{0}C), γ is the psychometric constant (k Pa^0 C^{-1}), R_n is the net solar radiation (MJm^{-2}hour^{-1}),), e_s is mean saturation vapour pressure(kPa), e_a is actual vapour pressure (kPa), G is the soil heat flux (MJm^{-2}hour^{-1}), T_m is the daily maximum

temperature (^{0}C), u_2 is the wind speed at 2m height (m/s), C_h is a constant with a value 0.24 during day time and 0.96 during night time.

To ensure faster learning and convergence, min-max normalization was applied to the input time series data for model training. Missing time series values were imputed using interpolation technique. Statistical properties of the processed data are listed in Table I.

The current forecasting problem is formulated as a univariate time-series forecasting task with in AutoGluon framework and it uses historical hourly ET_0 observations without incorporating exogenous meteorological variables. Hence, the time series data pre-processing tasks also includes standardisation and chronological arrangement of timestamp information, prior to the model development. The dataset was also converted into the TimeSeries DataFrame format within the AutoGluon framework. Chronological train–test splitting was subsequently performed by allocating 80% of the observations for model training and the remaining 20% for validation. This approach preserved temporal causality and prevented information leakage during forecasting experiments.

Table I: Descriptive statistics of meteorological variables of ACARR dataset

Parameter	**Maximum**	**Minimum**	**Mean**	**Standard deviation**
T_m	39.2	8.1	27.4	3.17
RH_{avg}	100	0.28	65.8	29.6
u_2	4.78	0.0	0.47	0.46
R_n	982.5	-149.5	130.7	232.3
ET_0	29.5	-4.1	3.95	6.55

Abbreviations: T_m: Average temperature (^{0}C), RH_{avg}: Average relative humidity (%), u_2: Wind speed (m/s), R_n: Net solar radiation (watts/m^2), ET_0: Reference evapotranspiration (mm/hour)

2.3 Forecasting Framework using AutoGluon-Timeseries

AutoGluon-Timeseries (AG-TS) (Shchur et al. 2023) is a recent open-source AutoML library used in developing the forecasting paradigms which can forecast the future values of multiple time series given the historical data and other related covariates. AG-TS offers both probabilistic and multi-horizon time-series forecasts as one unified forecasting environment by utilizing statistical forecasting models as well as other methods such as ML, DL architectures, and pretrained time series foundation models within a common benchmarking framework(Shchur et al. 2023).

AG-TS framework was primarily employed for automating the tedious tasks of training models, optimizing hyperparameters, and making forecasting(Westergaard et al. 2024). In this study, the forecasting experiments were implemented using the AG-TS framework and it solely based on historical hourly ET_0 values. It is a univariate forecasting framework integrating statistical, machine-learning, deep-learning, and ensemble forecasting paradigms(Shchur et al. 2023). Moreover, a comparative evaluation across multiple operational forecasting horizons was also be performed using AG-TS. Thus, it provides a means of comparing different types of forecasting methodologies in an efficient manner without requiring a lot of manual work to design and tune each model, which can be very

time-consuming. The ability to quickly compare various forecasting methods is especially beneficial for operational forecasting since scalability, reproducibility, and deployment efficiency are all important elements in any operational forecasting application(Sreedhara and Vanschoren 2024).

Furthermore, AG-TS automatically generates ensemble forecasting models by combining predictions from multiple trained models(Shchur et al. 2023). These ensemble approaches were also considered during comparative benchmarking to evaluate whether aggregated forecasting strategies could improve predictive robustness across different operational forecasting horizons. Fig. I shows the work flow diagram of AG-TS.

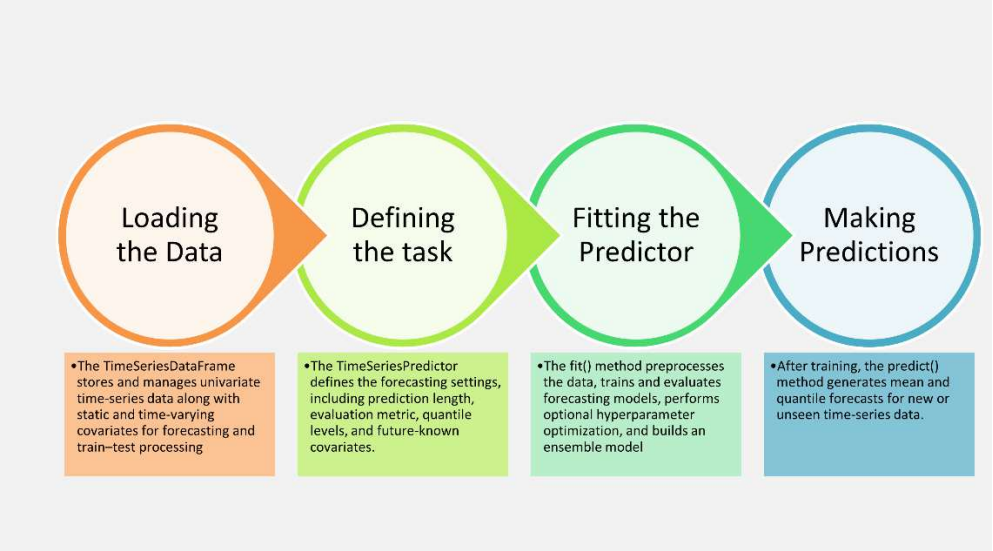


Fig. I Work flow diagram of AG-TS

2.4 Forecasting models under AG-TS

AG-TS combines various state of the art forecasting algorithms including classical statistical forecasting benchmarks such as SeasonalNaive, ETS, and Theta, attention-based deep learning models like DeepAR, transformer-based deep-learning forecasting such as Temporal Fusion Transformer (TFT), AutoGluon-Tabular forecasting strategies such as RecursiveTabular and DirectTabular, and a pretrained zero-shot forecasting model, Chronos2(Shchur et al. 2023).

2.4.1 Chronos models

Chronos models are the foundational models that have been built using a large and diverse collection of both real and synthetic time series datasets and can produce accurate out-of-the-box forecasts for unseen data(Ansari et al. 2024). There are various members of the Chronos family which are included in AG-TS. For example, chronos-t5-large, chronos-bolt-base, and Chronos2(Asher Marconi 2025). All of these models are appropriate for univariate forecasts; however, Chronos2 is pretrained model, particularly capable of cross learning across items and produces high-performance multivariate forecasts. Chronos2 extends

pretrained time-series forecasting by enabling univariate, multivariate, and covariate-informed forecasting within a unified zero-shot framework (Ansari et al. 2025). While do not really fit time series data, they can still produce forecasts through zero-shot inference based on contextual information provided to them. Chronos2 was built to allow for generalized sequence forecasting across a wide variety of temporal domains; therefore, this model uses large- scale temporal pre-training to develop transferable temporal representations that are suitable for supporting forecasting tasks with minimal manual feature engineering(Ansari et al. 2025). In the present study, Chronos2 was evaluated as the primary foundation-model architecture for short-term hourly ET_0 forecasting.

2.4.2 Temporal Fusion Transformer (TFT)

TFT is an attention-based deep-learning architecture designed for interpretable multi-horizon forecasting(Lim et al. 2021). Multi-horizon forecasting problems often contain a complex mix of inputs with no knowledge of how to interact with target variables. There are many DL models that have been developed to solve multi-horizon forecasting problems, but all have been developed as black box models that are not interpretable. However, the TFT provides high-performing multi-horizon forecasting along with interpretable insights on temporal dynamics. To learn temporal relationships at different scales, TFT integrates recurrent sequence modelling with attention mechanisms(Lim et al. 2021). Thus, it captures both short-term and long-term temporal dependencies. In addition, TFT employs specialized components to judiciously choose and select the features that are relevant and has a series of gating mechanism layers to suppress non-significant contributions, allowing the model to perform well over a broad variety of regimes(Rasiya Koya and Roy 2024). Due to its strong forecasting capability and increasing adoption in modern forecasting applications, TFT was included as a representative transformer-based benchmark model.

2.4.3 Tabular Forecasting Models

RecursiveTabular is a built-in automated tabular forecasting model that performs recursive future forecasting on time series data(Shchur et al. 2023); this means that each time step's future forecast will use the previous forecasted values as input to generate the next time step's forecast. Using a regression model developed as part of the AutoGluon-Tabular framework, RecursiveTabular makes future predictions one-at-a-time (Sreedhara and Vanschoren 2024). This recursive-based forecasting approach enables multi-step predictions with low computational cost. The model was included to evaluate the effectiveness of AutoML-assisted tabular forecasting for operational ET_0 prediction.

The DirectTabular model estimates future target variable values for each forecasting horizon without recursive dependence upon intermediate forecasts. Therefore, DirectTabular produces all future time series values simultaneously via an AutoGluon-Tabular regression model(Shchur et al. 2023). Direct forecasting generally propagates less cumulative error than recursive-based forecasting strategies over longer forecasting horizons. The model was included to compare direct and recursive forecasting behaviours within the AGT framework.

2.4.4 Statistical forecasting models

SeasonalNaive is a classical baseline forecasting model that predicts future values by replicating the last observed value from the same season(Makridakis et al. 2020). As ET_0 values fluctuate significantly depending on the time of the year, it is a highly seasonal

parameter. Moreover, Seasonal Naive provides a strong baseline performance in time-series forecasting tasks and is widely used as a benchmark model in forecasting studies and forecasting competitions.

The ETS model (Error, Trend and Seasonality) (Hyndman et al. 2002) is a family of forecasting methods that utilise the dynamics of time series by combining the trend, seasonal and error components in order to produce reliable forecasts using exponential smoothing to weight observations from the past in decreasing importance to each subsequent observation. The ETS method is very effective for producing forecasts for time series that demonstrate strong trend and seasonal characteristics. The ETS method also has the ability to account for time series data of different frequencies (hourly, daily, weekly, monthly and quarterly), which affects the way that seasons manifest themselves.

The Theta forecasting model (Assimakopoulos and Nikolopoulos 2000) is known for its strong forecasting ability in forecasting competitions and in operational forecasting applications. The Theta model uses a symmetrical approach that allows for the decomposition of time series into multiple theta lines in order to provide a more stable and accurate representation of trends within a time series.

Statistical forecasting models, including SeasonalNaive, ETS, and Theta, were incorporated as baseline forecasting paradigms within the AGT framework to provide benchmark comparisons against modern tabular, ensemble, deep-learning, and foundation forecasting approaches.

2.4.5 WeightedEnsemble

The AG-TS framework automatically creates an ensemble forecast using the WeightedEnsemble methodology to combine predictions obtained from multiple forecasting models that were trained independently(Shchur et al. 2023). By combining (via weighted averages) the strengths of different forecasting models, the ensemble forecast attempts to achieve superior prediction capability than any individual model. In the present study, the ensemble forecasting performance was evaluated alongside individual forecasting models across all forecasting horizons.

2.5 Multi-Horizon Forecasting Setup

The present study formulates the forecasting problem as a multi-horizon operational forecasting task. It evaluates model performance across different irrigation-planning timescales. Four forecasting horizons, namely 1 hour (1-h), 6 hours (6-h), 24 hours (24-h), and 168 hours (168-h), were selected to represent immediate, intraday, daily, and weekly operational forecasting scenarios, respectively.

The 1-h forecasting horizon allows for near real-time forecasting suitable for immediate irrigation-response applications. The 6-h horizon represents intraday irrigation scheduling under rapidly changing atmospheric conditions. The 24-h forecast horizon allows farmers to plan irrigation one day in advance and to plan in advance how much water they will have for distribution. Lastly, the 168-h horizon represents weekly operational forecasting relevant to medium-term agricultural water-resource management.

AG-TS framework executes separate forecasting experiments for each forecasting horizon by applying independent prediction lengths for each operational scenario. This methodology

helps to evaluate forecasting robustness and model performance incrementally across the progressively longer forecasting horizons. In general, forecasting uncertainty and temporal complexity increase with horizon length due to error propagation, nonlinear atmospheric variability, and reduced temporal dependency between historical and future observations.

The multi-horizon evaluation framework being used provides a comprehensive assessment of forecasting capability across multiple relevant operational forecasting timeframes rather than restricting evaluation to a single forecasting interval. This experimental design further enables comparative analysis of forecasting degradation patterns, robustness of pretrained foundation models, and horizon-wise performance variation among statistical, tabular, deep-learning, and ensemble forecasting approaches.

2.6 Model Evaluation Criteria

As this study evaluates the developed model's operational applicability across various forecasting horizons, prediction accuracy and robustness, there is a need of multiple statistical computational metrics for their performance assessment.

Mean absolute scaled error (MASE) was used as the primary ranking metric of this forecasting framework because of its scale-independent characteristics. Also, it is more suitable for comparative forecasting evaluation across different forecasting horizons. MASE provides normalized forecasting error estimates relative to naïve forecasting performance and is widely adopted in modern forecasting studies and forecasting competitions(Ansari et al. 2024).

Mean absolute error (MAE) and root mean square error (RMSE) were also employed to calculate the forecasting deviations between predicted and observed ET_0 values. RMSE assigns greater penalty to large forecasting errors. This helps to evaluate the forecasting stability under extreme variations. MAE provides a direct measure of average forecasting deviation with improved interpretability for operational applications (Mangalath Ravindran et al. 2022).

In addition to the forecasting accuracy, computational efficiency was also evaluated using model training and prediction times. Runtime evaluation is particularly important for operational agricultural forecasting systems where rapid forecasting generation and scalable deployment are desirable. All evaluation metrics were computed separately for each forecasting horizon to facilitate horizon-wise comparative analysis of forecasting robustness and model degradation behaviour.

2.6 Experimental Workflow of the proposed framework

Fig. II illustrates the overall workflow of the proposed AutoML-assisted multi-horizon ET_0 forecasting framework implemented using the AG-TS platform. Initially, hourly ET_0 data, calculated from meteorological observations using the ASCE Penman–Monteith equation, undergo preprocessing, including missing value imputation, timestamp standardization, and TimeSeriesDataFrame construction. Subsequently, the prepared ET_0 time series is configured for multi-horizon forecasting through a chronological 80:20 train–test split and prediction horizons of 1-h, 6-h, 24-h, and 168-h. The processed data are then supplied to the AG-TS framework, which automatically trains and evaluates multiple forecasting models, including

statistical, tabular, deep learning, pretrained foundation, and ensemble models. Finally, forecasts are generated and assessed using multiple performance metrics, followed by comparative benchmarking to identify the most suitable models for different forecasting horizons.

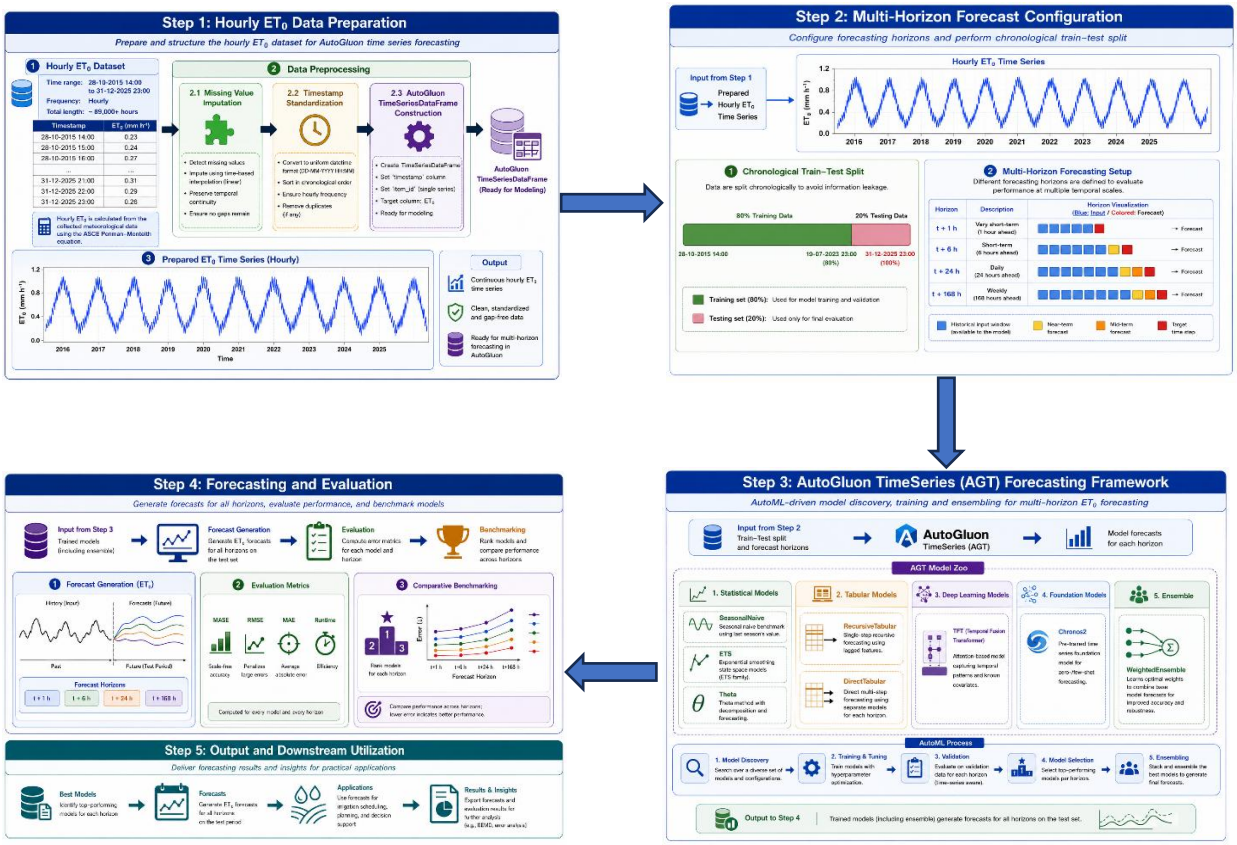


Fig. II Workflow diagram of the proposed AG-TS multi-horizon ET_0 forecasting framework

3. Results and Discussion

3.1 Forecasting Performance of AG-TS models across Operational Horizons

The forecasting performance of the evaluated models varied considerably across the investigated operational forecasting horizons, indicating horizon-dependent forecasting behaviour among different forecasting paradigms. Comparative benchmarking was performed using statistical, tabular, ensemble, deep-learning, and pretrained foundation forecasting models within the AG-TS framework. Tables II-V are the leader boards provided by AG-TS that displays each of the models it has trained, their scores on test and validation data, and training time in seconds, in decreasing order of their score. Internally, AG-TS anticipates a greater value for evaluation metrics during training, but the lowest value is really preferable in regression problems(Mangalath Ravindran et al. 2022). Hence, it reverses the sign of the metrics and reports them as negative values during training.

The obtained results demonstrated that no single forecasting model consistently dominated across all forecasting horizons. Instead, different forecasting paradigms exhibited varying strengths depending on the operational forecasting interval considered. For the 1-h forecasting

horizon, Chronos2 achieved the best forecasting performance with an MASE value of 0.096 among the evaluated models under the final experimental configuration. In contrast, WeightedEnsemble demonstrated comparatively superior performance for the 6-h and 24-h forecasting horizons with an MASE value of 0.8 and 0.647, respectively. At the 168-h forecasting horizon, Chronos2 again exhibited the best predictive performance among the evaluated models with an MASE value of 0.723. The score_test parameter in the leader boards showed a decrease in forecasting performance as the horizon length increases from 1 to 168. By observing an 86.7% drop in MASE values (0.096 to 0.723) from 1-h to 168-h, the table entries establish the significance of horizon length in the improved ET_0 forecast accuracy. However, within the same horizon, there exists 3–53% decrease in performance variation among top three leading models. The graph shown in Fig. III demonstrates the horizon-wise forecasting performance comparison of the evaluated forecasting models.

Table VI presents the forecasting performance of the AG-TS leader models across multiple operational forecasting horizons during the testing period. The obtained results demonstrate horizon-dependent variations in forecasting behaviour, consistent with the patterns observed during leaderboard-based model evaluation. At the 1-h forecasting horizon, the lowest MAE, RMSE, and MASE values were obtained, indicating the strong short-term forecasting capability of Chronos2. Among the remaining forecasting horizons, the 24-h horizon exhibited comparatively lower forecasting errors than the 6-h and 168-h horizons. This behaviour may be attributed to the pronounced diurnal periodicity and recurring atmospheric patterns governing hourly ET_0 dynamics, which enhance predictability at the day-ahead forecasting horizon.

These findings indicate that forecasting performance is strongly influenced by forecasting horizon and underlying temporal complexity. Short-horizon forecasting primarily depends on capturing immediate temporal continuity and local sequential patterns, whereas extended forecasting horizons require modelling of broader temporal dependencies and cumulative forecasting uncertainty. Overall, the results demonstrate the effectiveness of the AG-TS framework for automated benchmarking of heterogeneous forecasting paradigms in hourly ET_0 prediction. Furthermore, the competitive performance of Chronos2 at both the shortest and largest forecasting horizons considered in the present study highlights the potential applicability of pretrained time-series foundation models for operational short-term hourly ET_0 forecasting under the investigated dataset conditions.

Table II: Leader board of AG-TS models on 1-h forecasting paradigm

Model	score_test	score_val	pred_time_test	pred_time_val	fit_time_marginal	fit_order
Chronos2	-0.096	-0.249	1.057	2.151	19.206	6
DirectTabular	-0.179	-0.439	1.290	0.800	7.217	3
WeightedEnsemble	-0.203	-0.533	0.213	0.194	0.370	8
Theta	-0.330	-0.035	0.265	0.054	0.058	5
TemporalFusionTransformer	-0.330	-0.035	0.208	0.053	474.034	7
ETS	-0.997	-1.414	2.444	3.226	0.052	4
RecursiveTabular	-1.817	-0.724	0.127	0.125	3.778	2
SeasonalNaive	-2.163	-0.332	0.016	0.010	0.062	1

Table III: Leader board of AG-TS models on 6-h forecasting paradigm

model	score_test	score_val	pred_time_test	pred_time_val	fit_time_marginal	fit_order
WeightedEnsemble	-0.800	-0.253	1.645	0.802	0.363	8
Chronos2	-0.815	-0.223	1.921	0.656	0.338	6
DirectTabular	-0.838	-0.189	1.447	0.856	6.145	3
TemporalFusionTransformer	-0.941	-0.190	0.192	0.052	816.157	7
SeasonalNaive	-1.192	-0.339	0.017	0.010	0.050	1
RecursiveTabular	-1.247	-0.304	0.231	0.130	3.654	2
Theta	-2.499	-0.330	0.333	0.263	0.054	5
ETS	-3.327	-0.239	8.761	1.960	0.056	4

Table IV: Leader board of AG-TS models on 24-h forecasting paradigm

model	score_test	score_val	pred_time_test	pred_time_val	fit_time_marginal	fit_order
WeightedEnsemble	-0.647	-0.712	0.248	0.819	0.581	8
DirectTabular	-0.664	-0.747	0.839	0.075	6.131	3
Chronos2	-0.668	-0.680	1.006	0.202	0.793	6

TemporalFusionTransformer	-0.668	-0.679	0.121	0.219	724.653	7
SeasonalNaive	-0.693	-0.694	0.018	0.014	0.075	1
RecursiveTabular	-0.843	-0.837	1.006	0.643	4.246	2
Theta	-2.600	-1.748	0.282	0.331	0.086	5
ETS	-9.188	-7.239	1.977	3.666	0.101	4

Table V: Leader board of AG-TS models on 168-h forecasting paradigm

model	score_test	score_val	pred_time_test	pred_time_val	fit_time_marginal	fit_order
Chronos2	-0.723	-1.573	1.050	0.663	0.337	6
WeightedEnsemble	-0.831	-1.193	3.018	1.953	0.402	8
TemporalFusionTransformer	-0.849	-1.214	0.217	1.102	2666.883	7
DirectTabular	-0.933	-1.226	1.567	0.091	16.633	3
RecursiveTabular	-1.053	-1.298	1.226	0.755	9.982	2
SeasonalNaive	-1.191	-1.389	0.024	0.026	0.073	1
Theta	-1.337	-3.264	0.375	0.283	0.055	5
ETS	-14.697	-21.296	2.108	8.357	0.080	4

Table VI: Comparative forecasting performance of AG-TS leader models on ACARR dataset across various forecasting horizons during testing period

Horizon	Model	MASE	RMSE	MAE	Training Time in seconds	Prediction Time in seconds
1-h	Chronos2	0.327	0.268	0.214	524.692	3.073312
6-h	WeightedEnsemble	0.830	0.901	0.680	831.9529	3.405457
24-h	WeightedEnsemble	0.661	0.833	0.541	743.8856	2.703206
168-h	Chronos2	0.822	1.005	0.672	2707.654	6.289483

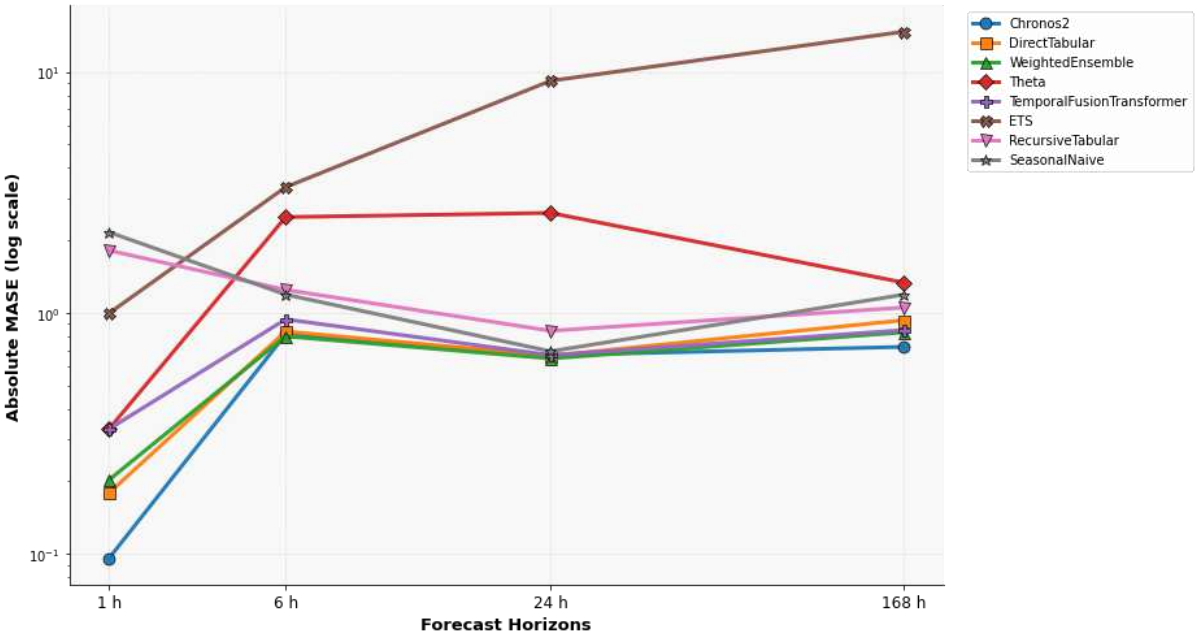


Fig. **III** Horizon-wise forecasting performance comparison of the evaluated forecasting models

3.2 Comparative Behaviour of Forecasting Paradigms

Table VII shows the best performing forecasting paradigm across the four operational forecasting horizons such as 1-h, 6-h, 24-h and 168-h. The comparative evaluation revealed distinct forecasting characteristics among the investigated forecasting paradigms, such as tabular, ensemble, deep learning, foundation and statistical forecasting. At 1-h forecasting, the ET_0 series is highly autocorrelated, temporally smooth, and dominated by immediate sequential continuity. The 6-h and 24-h forecasting is often influenced strongly by recent lag relationships, local periodicity, intraday temporal structures, and deterministic hourly patterns. And in 168-h weekly horizon, forecasting becomes more uncertain, multi-step accumulated, less dependent on immediate lags, and more dependent on broader temporal representations.

The pre-trained foundation model, Chronos2 demonstrated comparatively strong performance at both the 1-h and 168-h forecasting horizons, with score_test values of 0.096 and 0.723, respectively. Moreover, an average improvement of 60% and 20% in forecasting performance was observed for Chronos2 in the 1-h and 168-h horizons, respectively when compared to other forecasting paradigms. Chronos2, being a pretrained foundation model, may effectively capture short-term temporal continuity, periodic sequence behaviour, and immediate temporal dependencies through its pretrained temporal representations. Furthermore, it may benefit from pretrained generalized temporal representations, long-context sequence learning, and temporal abstraction capabilities, thereby supporting improved forecasting performance across extended horizons. Hence, Chronos2 demonstrated strong forecasting capability at both the immediate (1-h) and extended (168-h) forecasting horizons under the investigated dataset conditions, thereby demonstrating its suitability for multi-step hourly ET_0 forecasting.

The ensemble forecasting paradigm, WeightedEnsemble exhibited superior forecasting performance for the 6-h and 24-h forecasting horizons with a score_test parameter of 0.8 and 0.657, respectively. In addition, in the other two horizons as well, the score_test values show only a slight variation of 0.1 compared to those of the best-performing models. Hence, the WeightedEnsemble model performs comparably to the best-performing model, Chronos2, in the 1-h and 168-h horizons, thereby demonstrating its forecasting efficiency across all the horizons considered. Thus, the comparatively superior performance of the WeightedEnsemble model at the 6-h and 24-h forecasting horizons suggests that ensemble-based forecasting may effectively capture the complex temporal behaviour of hourly ET_0 at intermediate operational horizons. Forecasting at these horizons is likely influenced by various factors such as recent lag relationships, local periodicity, intraday atmospheric variability, and short-term nonlinear dependencies, which may not be fully represented by a single forecasting paradigm. By integrating predictions from multiple constituent models, the WeightedEnsemble approach may benefit from complementary forecasting strengths while reducing model-specific forecasting uncertainty under the investigated dataset conditions.

Among the tabular forecasting paradigms, DirectTabular demonstrated competitive performance, particularly for short- and intermediate-horizon forecasting, where it ranked among the top three best-performing models. DirectTabular achieved strong forecasting performance, with score_test values ranging from 0.178 to 0.66 across the 1-h, 6-h, and 24-h forecasting horizons, which were very close to those of the best-performing Chronos2 and WeightedEnsemble models. This indicates the effectiveness of lag-based tabular learning for capturing short-term temporal dependencies within hourly ET_0 series. The strong performance of DirectTabular suggests that local temporal continuity and immediate historical patterns play important roles in immediate, intraday, and daily ET_0 forecasting.

Statistical forecasting paradigms such as SeasonalNaive, ETS, and Theta generally exhibited lower predictive performance when compared with modern machine-learning and foundation-model approaches, particularly at longer forecasting horizons, 24-h and 168-h. The score_test parameter of these statistical models ranges from 0.33 to 14.69, indicating comparatively lower forecasting performance than tabular, ensemble, deep learning, and pretrained foundation forecasting models, whose score_test values range from 0.09 to 1.24. Although these models provided computationally efficient forecasting, their ability to capture complex nonlinear temporal dynamics within hourly ET_0 observations was comparatively limited.

Table VII: Best-performing forecasting paradigm across operational forecasting horizons

Horizon	**Best Model**	**Forecasting Paradigm**
1-h	Chronos2	Pre-trained foundation Model
6-h	WeightedEnsemble	Ensemble Model
24-h	WeightedEnsemble	Ensemble Model
168-h	Chronos2	Pre-trained foundation Model

3.3 Horizon-Wise Forecasting Characteristics

Forecasting uncertainty generally increased with increasing forecasting horizon, reflecting the growing complexity of multi-step ET_0 prediction. The 1-h forecasting horizon primarily

involved immediate temporal continuity and short-term persistence patterns, whereas the 168-h forecasting horizon required modelling broader sequential dependencies and accumulated forecasting uncertainty over extended prediction intervals. Fig. IV demonstrates the actual and predicted hourly ET_0 values for selected forecasting horizons using the best-performing forecasting models.

The obtained results indicate that different forecasting paradigms exhibit varying sensitivity to forecasting horizon. Ensemble forecasting approaches provided comparatively stable performance at intermediate forecasting horizons by integrating predictions from multiple forecasting paradigms. In contrast, Chronos2 demonstrated comparatively robust performance for both immediate and weekly forecasting horizons under the investigated experimental conditions. Tabular forecasting approaches demonstrated strong capability for short-term operational forecasting, likely due to their ability to effectively exploit local lag-based information.

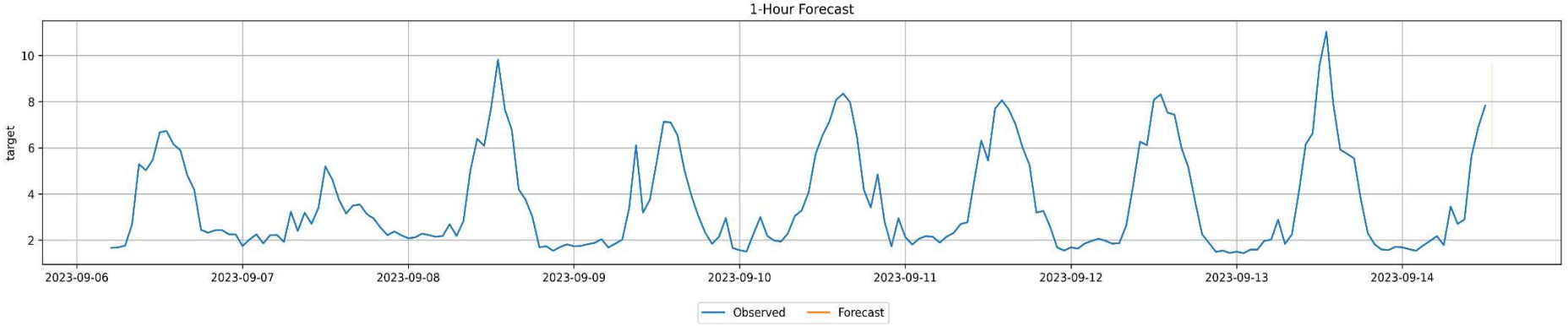
The horizon-dependent variation observed in the present study further highlights the importance of evaluating forecasting models across multiple operational forecasting intervals rather than relying on a single prediction horizon. Such multi-horizon benchmarking provides improved understanding of forecasting robustness, model stability, and operational applicability under realistic agricultural forecasting conditions.

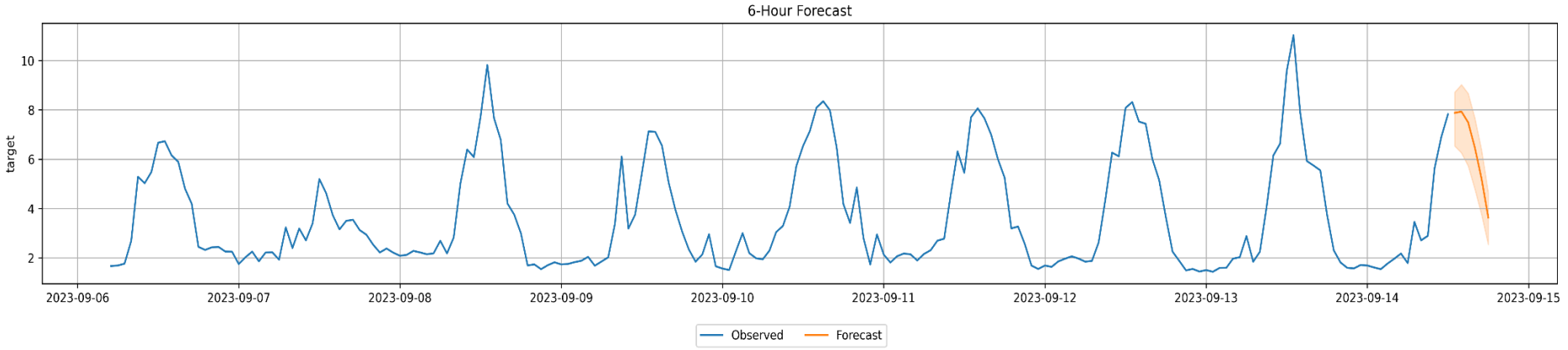
3.4 Computational Efficiency and Operational Applicability

Besides forecasting accuracy, computational efficiency represents an important consideration for operational agricultural forecasting systems. Table VIII demonstrated that the evaluated forecasting models exhibited considerable variation in training and prediction runtimes. Statistical forecasting models generally required minimal computational resources, whereas deep-learning approaches exhibited comparatively higher computational demand. The line graphs shown in Fig. V provides a comparative analysis of the computational efficiency of forecasting models.

From a computational perspective, WeightedEnsemble demonstrated moderate training (0.43 s) and prediction (1.28 s) runtimes, indicating a favourable balance between forecasting performance and computational demand. Despite relying on multiple underlying forecasting models, the comparatively low computational cost observed in the present study suggests its practical suitability for operational intraday and day-ahead ET_0 forecasting systems.

Despite being a pretrained foundation model, Chronos2 demonstrated relatively low training (5.17 s) and prediction (1.26 s) runtimes while maintaining competitive forecasting performance across multiple forecasting horizons. This characteristic may provide advantages for operational forecasting applications where rapid model deployment and forecasting generation are desirable. The comparatively efficient runtime behaviour observed for Chronos2, especially when compared with computationally intensive deep-learning approaches such as TFT, suggests that pretrained forecasting architectures may reduce the need for extensive model-specific optimization and training procedures.


1-Hour Forecast
target
10
8
6
4
2
2023-09-06
2023-09-07
2023-09-08
2023-09-09
2023-09-10
2023-09-11
2023-09-12
2023-09-13
2023-09-14
Observed
Forecast


6-Hour Forecast
target
10
8
6
4
2
2023-09-06
2023-09-07
2023-09-08
2023-09-09
2023-09-10
2023-09-11
2023-09-12
2023-09-13
2023-09-14
2023-09-15
Observed
Forecast

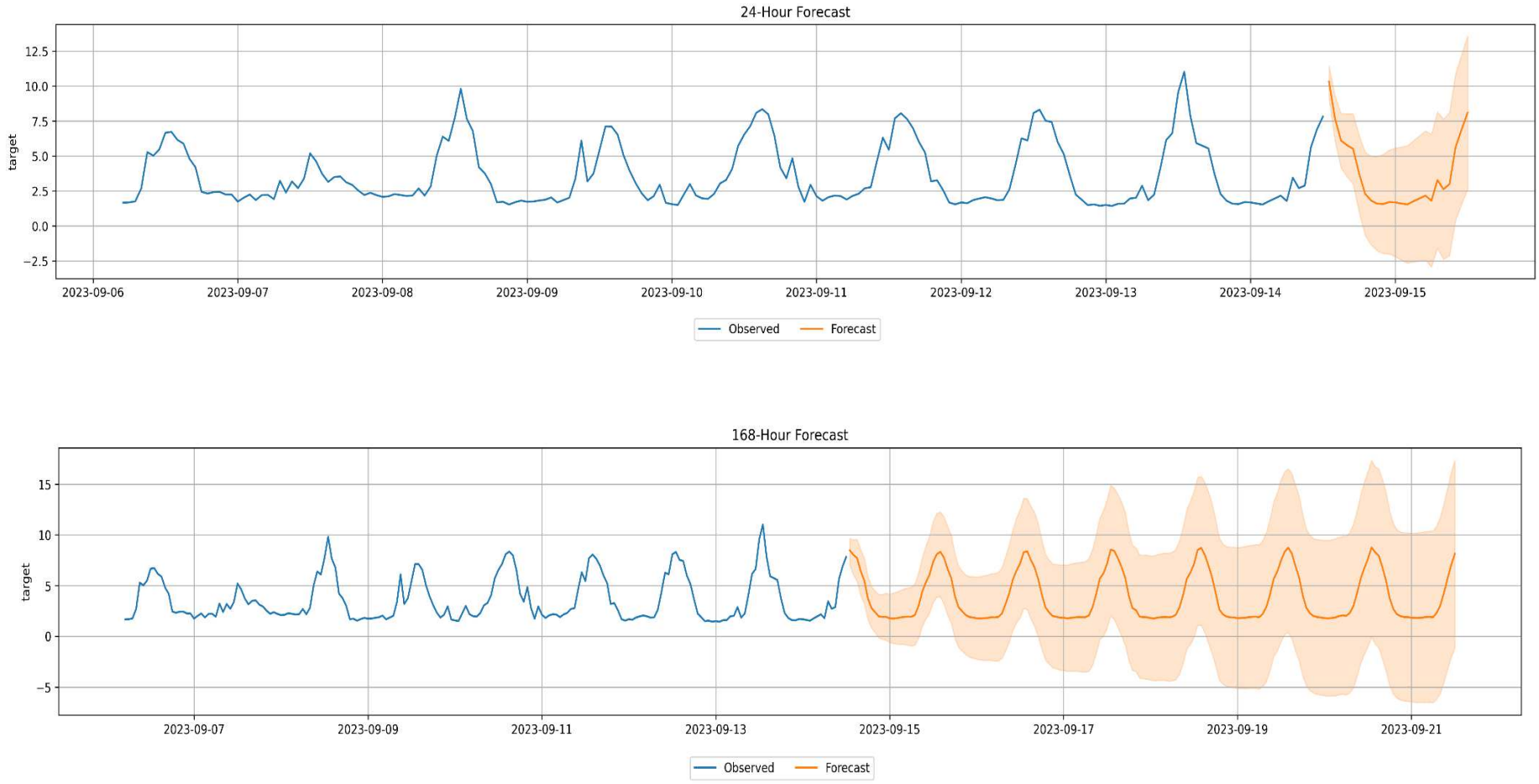


Fig. IV Actual and predicted hourly ET_0 values for selected forecasting horizons using the best-performing forecasting model

From an operational perspective, both tabular forecasting approaches demonstrated favourable computational efficiency compared with deep-learning models, suggesting their practical applicability in forecasting systems requiring relatively low computational overhead and rapid model deployment.

TFT demonstrated substantially higher computational demand compared with other evaluated forecasting paradigms, requiring the largest average training time among all investigated models. Although TFT exhibited rapid prediction runtime after training, its extensive computational requirement may limit operational applicability in environments with constrained computational resources or requirements for rapid model updating.

These findings indicate that computationally intensive deep-learning architectures may not always provide operational advantages for hourly ET_0 forecasting, particularly when comparatively efficient forecasting paradigms such as foundation, ensemble, and tabular models achieve competitive or superior forecasting performance.

Table VIII: Average computational efficiency and operational applicability of evaluated forecasting models across operational forecasting horizons

Model	Forecasting Paradigm	Avg. Training Time (s)	Avg. Prediction Time (s)	Operational implication
Theta	Statistical forecasting	0.06	0.31	Simple and computationally efficient baseline suitable for rapid forecasting
ETS	Statistical forecasting	0.07	3.82	Lightweight statistical forecasting approach with low training requirements
SeasonalNaive	Statistical forecasting	0.07	0.02	Minimal computational cost, suitable as a benchmark forecasting baseline
WeightedEnsemble	Ensemble forecasting	0.43	1.28	Robust forecasting through aggregation of multiple models, though dependent on constituent model outputs.
Chronos2	Foundation model	5.17	1.26	Competitive forecasting with comparatively low computational demand, suitable for operational multi-horizon forecasting.
Recursive Tabular	Tabular forecasting	5.42	0.65	Efficient lag-based forecasting with suitability for short- and intermediate-horizon prediction; recursive forecasting may accumulate uncertainty over longer horizons.
DirectTabular	Tabular forecasting	9.03	1.29	Effective for structured lag-based forecasting and intraday temporal dependency modelling.
TFT	Deep learning	1170.43	0.18	High computational complexity with substantial training requirements despite rapid inference capability.

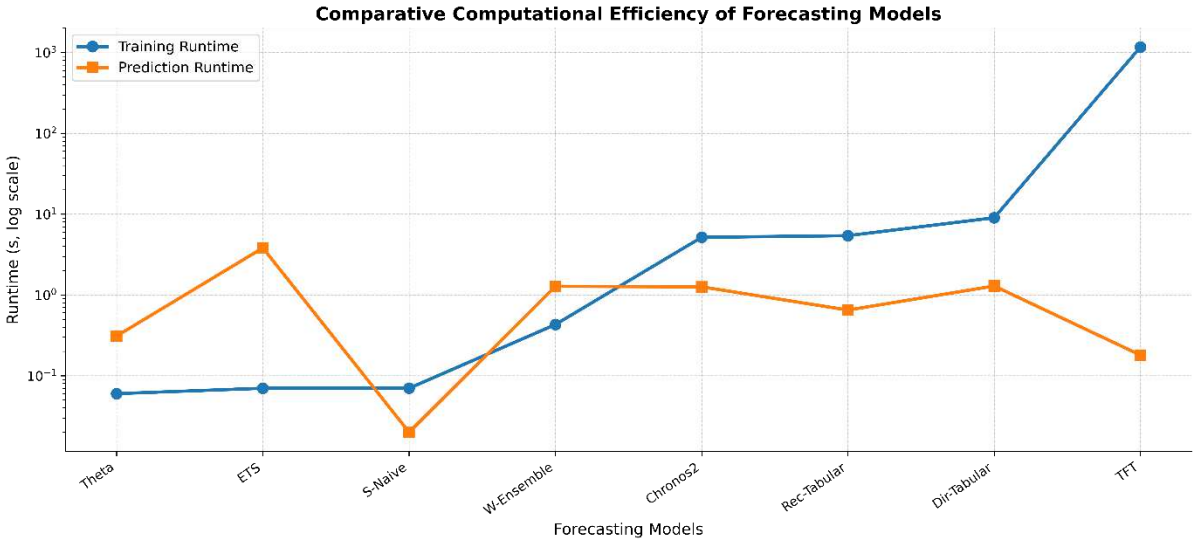


Fig. V Line graphs comparing computational efficiency of forecasting models

Fig. VI and Fig. VII illustrate the training and prediction runtime distribution across the four forecast horizons to assess the computational efficiency and operational applicability of the various forecasting paradigms.

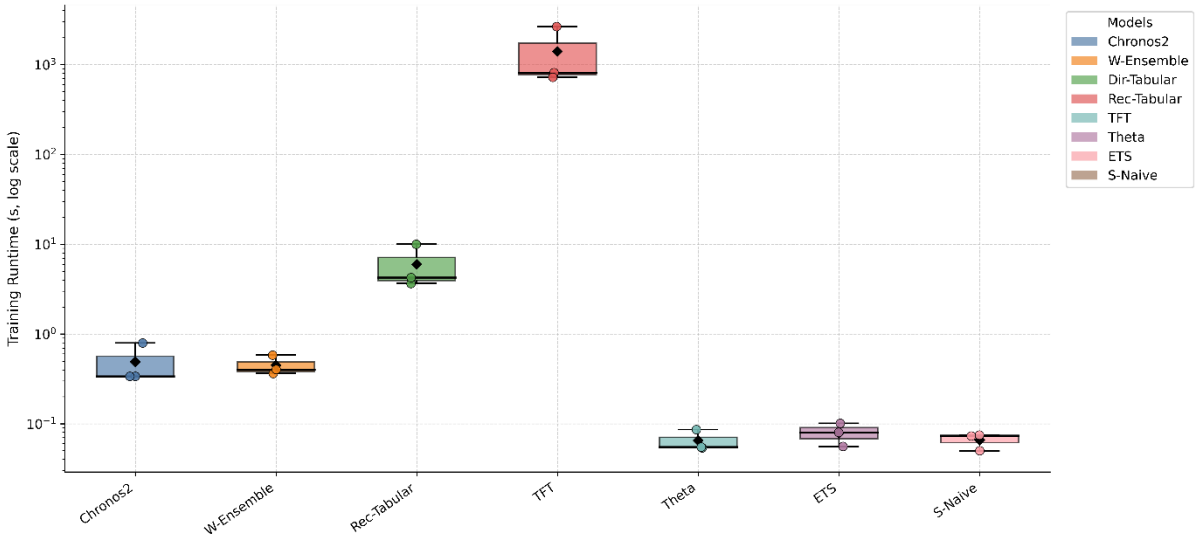


Fig. VI Box plot showing Training Runtime Distribution across Forecast Horizons

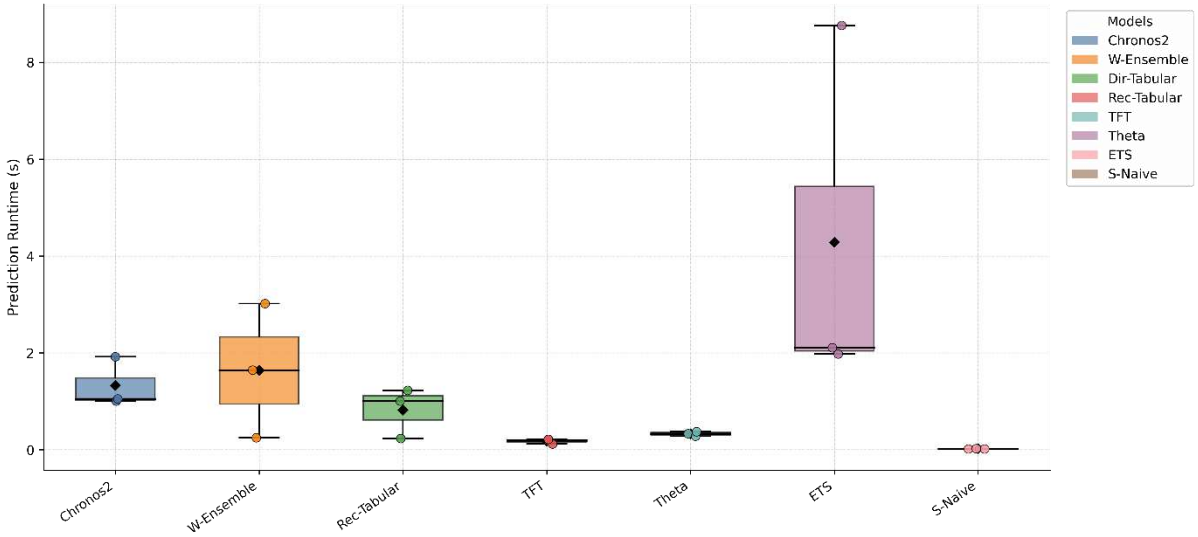


Fig. VII Box plot showing Prediction Runtime Distribution across Forecast Horizons

5. Conclusion

This study presented an automated multi-horizon forecasting framework for hourly ET_0 forecasting using statistical, tabular, ensemble, deep-learning, and foundation forecasting paradigms implemented within the AG-TS framework. Hourly ET_0 values derived using the ASCE Penman–Monteith equation from meteorological observations spanning October 2015 to December 2025 were forecasted across four operational horizons (1-h, 6-h, 24-h, and 168-h). The forecasting performance of AG-TS built-in models such as SeasonalNaive, Theta, ETS, RecursiveTabular, DirectTabular, WeightedEnsemble, Temporal Fusion Transformer (TFT), and Chronos2 models was comparatively evaluated.

The findings revealed clear horizon-dependent forecasting behaviour. Chronos2 demonstrated superior predictive capability at the immediate (1-h) and weekly (168-h) forecasting horizons, indicating the potential capability of pretrained time-series foundation models for capturing both short-term temporal continuity and extended multi-step dependencies in hourly ET_0 forecasting. In contrast, WeightedEnsemble achieved superior performance at the 6-h and 24-h forecasting horizons, suggesting that ensemble integration of complementary forecasting paradigms may be particularly beneficial for intermediate operational forecasting. Tabular forecasting approaches exhibited competitive performance with comparatively moderate computational requirements, whereas TFT incurred substantially higher training costs despite rapid inference capability. Statistical forecasting models showed the least forecasting performance, despite their minimal computational resource requirement.

From an operational perspective, the obtained results demonstrate the potential of automated forecasting frameworks for robust short-term ET_0 prediction to support irrigation scheduling and water-resources planning. Although the findings are dataset-specific, they highlight the emerging applicability of pretrained foundation forecasting models in agro-hydrological contexts. Future research may investigate longer observational records, additional hydro-agro-climatic settings, exogenous-variable integration, and explainable artificial intelligence approaches to further improve model interpretability and generalizability.

Acknowledgements

The author would like to express their deep gratitude to the Advanced Centre for Atmospheric Radar Research (ACARR) in Cochin University of Science and Technology, Kochi, Kerala, India, supported by the Ministry of Earth Sciences (MoES), Govt. of India, for providing meteorological data for this study.

References

Afifi A (2020) Forecasting Reference Evapotranspiration Using Non-Linear Autoregressive Artificial Neural Network. Master's thesis, University of Hawai'i at Manoa.

Aghelpour P, Norooz-Valashedi R (2022) Predicting daily reference evapotranspiration rates in a humid region, comparison of seven various data-based predictor models. Stoch Environ Res Risk Assess 36:4133–4155. https://doi.org/10.1007/s00477-022-02249-4

Ahmed AAM, Deo RC, Feng Q, et al (2022) Hybrid deep learning method for a week-ahead evapotranspiration forecasting. Stoch Environ Res Risk Assess 36:831–849. https://doi.org/10.1007/s00477-021-02078-x

Alibabaei K, Gaspar PD, Lima TM (2020) Modeling evapotranspiration using Encoder-Decoder Model. 2020 Int Conf Decis Aid Sci Appl DASA 2020 132–136. https://doi.org/10.1109/DASA51403.2020.9317100

Allen RG, Pruitt WO, Wright JL, et al (2006) A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. Agric Water Manag 81:1–22. https://doi.org/10.1016/J.AGWAT.2005.03.007

Ansari AF, Shchur O, Küken J, et al (2025) Chronos-2: From Univariate to Universal Forecasting. arXiv preprint arXiv:2510.15821.

Ansari AF, Stella L, Turkmen C, et al (2024) Chronos: Learning the Language of Time Series. Trans Mach Learn Res 2024. arXiv preprint arXiv:2403.07815

Asher Marconi B (2025) Time Series Foundation Models for Multivariate Financial Time Series Forecasting. https://doi.org/10.2139/SSRN.6085266

Assimakopoulos V, Nikolopoulos K (2000) The theta model: a decomposition approach to forecasting. Int J Forecast 16:521–530. https://doi.org/10.1016/S0169-2070(00)00066-2

Babaeian E, Paheding S, Siddique N, et al (2022) Short- and mid-term forecasts of actual evapotranspiration with deep learning. J Hydrol 612:128078. https://doi.org/10.1016/J.JHYDROL.2022.128078

Bachour R, Maslova I, Ticlavilca AM, et al (2016) Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. Stoch Environ Res risk Assess 30:103–117. https://doi.org/10.1007/s00477-015-1039-z

Bedi J (2022) Transfer learning augmented enhanced memory network models for reference evapotranspiration estimation. Knowledge-Based Syst 237:107717. https://doi.org/10.1016/J.KNOSYS.2021.107717

Chen Y, Liu Z, Long T, et al (2025) Evaluation of Eight Decomposition-Hybrid Models for Short-Term Daily Reference Evapotranspiration Prediction. Atmosphere (Basel) 16:. https://doi.org/10.3390/atmos16050535

Chia MY, Huang YF, Koo CH, Fung KF (2020) Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—A review. Agronomy 10:. https://doi.org/10.3390/agronomy10010101

de Oliveira e Lucas P, Alves MA, de Lima e Silva PC, Guimarães FG (2020) Reference evapotranspiration time series forecasting with ensemble of convolutional neural networks. Comput Electron Agric 177:105700. https://doi.org/10.1016/j.compag.2020.105700

Ferreira LB, da Cunha FF (2020a) Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. Comput Electron Agric 178:105728. https://doi.org/10.1016/J.COMPAG.2020.105728

Ferreira LB, da Cunha FF (2020b) New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. Agric Water Manag 234:106113. https://doi.org/10.1016/j.agwat.2020.106113

Gao Z, Gao Z, Sun C, et al (2026) Integrating Seasonal-Trend Decomposition and Multi-Head Self-Attention for Feature Fusion with Bidirectional Long Short-Term Memory in Reference Evapotranspiration Prediction. Water Resour Manag 40:. https://doi.org/10.1007/s11269-026-04553-y

Gao Z, Yang D, Li B, et al (2025) Reference Crop Evapotranspiration Prediction Based on Gated Recurrent Unit with Quantum Inspired Multi-head Self-attention Mechanism. Water Resour Manag 39:1481–1501. https://doi.org/10.1007/s11269-024-04016-2

Goyal P, Kumar S, Sharda R (2023) A review of the Artificial Intelligence (AI) based techniques for estimating reference evapotranspiration: Current trends and future perspectives. Comput Electron Agric 209:107836. https://doi.org/10.1016/J.COMPAG.2023.107836

Gupta S, Bandyopadhyay KK, Sarangi A, et al (2026) Reference evapotranspiration estimation using AI: performance evaluation across different climates – a review. Theor Appl Climatol 2026 1572 157:77-. https://doi.org/10.1007/S00704-025-05934-W

Hyndman RJ, Koehler AB, Snyder RD, Grose S (2002) A state space framework for automatic forecasting using exponential smoothing methods. Int J Forecast 18:439–454. https://doi.org/10.1016/S0169-2070(01)00110-8

Karbasi M (2018) Forecasting of multi-step ahead reference evapotranspiration using wavelet-Gaussian process regression model. Water Resour Manag 32:1035–1052. https://doi.org/10.1007/s11269-017-1853-9

Karbasi M, Ali M, Randhawa GS, et al (2024) Innovative multi-temporal evapotranspiration forecasting using empirical fourier decomposition and bidirectional long short-term memory. Smart Agric Technol 9:. https://doi.org/10.1016/j.atech.2024.100619

Karbasi M, Jamei M, Ali M, et al (2023) Development of an enhanced bidirectional recurrent neural network combined with time-varying filter-based empirical mode decomposition to forecast weekly reference evapotranspiration. Agric Water Manag 290:108604. https://doi.org/10.1016/j.agwat.2023.108604

Landeras G, Ortiz-Barredo A, López JJ (2009) Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. J Irrig Drain Eng 135:323–334. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000008

Lee J, Bateni SM, Jun C, et al (2024) Hybrid machine learning system based on multivariate data decomposition and feature selection for improved multitemporal evapotranspiration

forecasting. Eng Appl Artif Intell 135:. https://doi.org/10.1016/j.engappai.2024.108744

Li M, Zhou Q, Han X, Lv P (2024) Prediction of reference crop evapotranspiration based on improved convolutional neural network (CNN) and long short-term memory network (LSTM) models in Northeast China. J Hydrol 645:. https://doi.org/10.1016/j.jhydrol.2024.132223

Lim B, Arık S, Loeff N, Pfister T (2021) Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. Int J Forecast 37:1748–1764. https://doi.org/10.1016/J.IJFORECAST.2021.03.012

Lu Y, Li T, Hu H, Zeng X (2023) Short-term prediction of reference crop evapotranspiration based on machine learning with different decomposition methods in arid areas of China. Agric Water Manag 279:108175. https://doi.org/10.1016/j.agwat.2023.108175

Makridakis S, Spiliotis E, Assimakopoulos V (2020) The M4 Competition: 100,000 time series and 61 forecasting methods. Int J Forecast 36:54–74. https://doi.org/10.1016/J.IJFORECAST.2019.04.014

Malik A, Jamei M, Ali M, et al (2022) Multi-step daily forecasting of reference evapotranspiration for different climates of India: A modern multivariate complementary technique reinforced with ridge regression feature selection. Agric Water Manag 272:107812. https://doi.org/https://doi.org/10.1016/j.agwat.2022.107812

Mangalath Ravindran S, Moorakkal Bhaskaran SK, K. Ambat S, et al (2022) An automated machine learning methodology for the improved prediction of reference evapotranspiration using minimal input parameters. Hydrol Process 36:e14571. https://doi.org/https://doi.org/10.1002/hyp.14571

Nourani V, Elkiran G, Abdullahi J (2020) Multi-step ahead modeling of reference evapotranspiration using a multi-model approach. J Hydrol 581:124434. https://doi.org/10.1016/j.jhydrol.2019.124434

Park YJ, Germain F, Liu J, et al (2025) Probabilistic forecasting for building energy systems using time-series foundation models. Energy Build 348:116446. https://doi.org/10.1016/J.ENBUILD.2025.116446

Proias G, Gravalos I, Papageorgiou E, et al (2020) Forecasting reference evapotranspiration using time lagged recurrent neural network. WSEAS Trans Environ Dev 16:699–707. https://doi.org/10.37394/232015.2020.16.72

Qian H, Wang W, Chen G (2024) Assessing forecast performance of daily reference evapotranspiration: A comparison of equations, machine and deep learning using weather forecasts. J Hydrol 644:. https://doi.org/10.1016/j.jhydrol.2024.132101

Rasiya Koya S, Roy T (2024) Temporal Fusion Transformers for streamflow Prediction: Value of combining attention with recurrence. J Hydrol 637:131301. https://doi.org/10.1016/J.JHYDROL.2024.131301

Sang YF (2013) A review on the applications of wavelet transform in hydrology time series analysis. Atmos Res 122:8–15. https://doi.org/10.1016/J.ATMOSRES.2012.11.003

Shchur O, Turkmen C, Erickson N, et al (2023) AutoGluon-TimeSeries: AutoML for Probabilistic Time Series Forecasting. Proc Mach Learn Res 228:

Sreedhara AT, Vanschoren J (2024) Can time series forecasting be automated? A benchmark and analysis. 1–38

Tang M, Ma Z, Ai P, et al (2025) Multi-step-ahead forecasting of daily reference evapotranspiration using hybrid deep learning models for the Taklamakan Desert oasis. J Hydrol Reg Stud 61:. https://doi.org/10.1016/j.ejrh.2025.102663

Torres AF, Walker WR, McKee M (2011) Forecasting daily potential evapotranspiration using machine learning and limited climatic data. Agric Water Manag 98:553–562

Wang S, Yu X, Li Y, et al (2023) Application of a hybrid deep learning approach with attention mechanism for evapotranspiration prediction: a case study from the Mount Tai region, China. Earth Sci Informatics 16:3469–3487. https://doi.org/10.1007/s12145-023-01103-7

Westergaard G, Erden U, Mateo OA, et al (2024) Time Series Forecasting Utilizing Automated Machine Learning (AutoML): A Comparative Analysis Study on Diverse Datasets. Inf 15:. https://doi.org/10.3390/info15010039

Willard JD, Varadharajan C, Jia X, Kumar V (2025) Time series predictions in unmonitored sites: a survey of machine learning techniques in water resources. Environ Data Sci 4:e7. https://doi.org/10.1017/EDS.2024.14

Xiang K, Li Y, Horton R, Feng H (2020) Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review. Agric Water Manag 232:106043. https://doi.org/10.1016/J.AGWAT.2020.106043

Yan Y, Li G, Li Q, et al (2026) Improving Short-term Forecasts of Soil Moisture and Evapotranspiration with Multi-task Learning and Transformer-based Feature Processing. Water Resour Manag 40:1–17. https://doi.org/10.1007/s11269-026-04591-6

Yan Z, Lu X, Wu L (2025) Evaluating the impact of different decomposition methods on the accuracy of reference evapotranspiration forecasts in humid regions. J Hydroinformatics 27:406–441. https://doi.org/10.2166/hydro.2025.170

Ye Y, González-Vidal A, Zamora-Izquierdo MA, Skarmeta AF (2025) Transfer and deep learning models for daily reference evapotranspiration estimation and forecasting in Spain from local to national scale. Smart Agric Technol 11:. https://doi.org/10.1016/j.atech.2025.100886

Zhang J, Ding Y, Zhu L, et al (2025) Estimating and forecasting daily reference crop evapotranspiration in China with temperature-driven deep learning models. Agric Water Manag 307:109268. https://doi.org/10.1016/j.agwat.2024.109268

Zhang X, Sridharan S, Zahrin NH Bin, et al (2026) Benchmarking a time-series foundation model (TimeGPT) for real-world forecasting applications. Mach Learn with Appl 23:100801. https://doi.org/10.1016/j.mlwa.2025.100801

Zhou H, Su Y, Ma L, et al (2026) Optimizing light gradient boosting machine with the slime mould algorithm for reference evapotranspiration estimation. Agric Water Manag 324:. https://doi.org/10.1016/j.agwat.2025.110107

Statements and Declarations

Funding

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Conceptualization, Methodology, Software, and Validation were performed by Sowmya M R. Investigation, Resources, Project administration, and Writing- Review were done by Santosh Kumar M B. All authors read and approved the final manuscript.