RESEARCH ARTICLE

Performance Evaluation of ETo Prediction Methods: Dispersion Analysis and Accuracy Criteria Across Time Intervals

Mostafa Sadeghzadeh[1,*], **Jalal Shiri**[1,2] and **Sepideh Karimi**[1]

[1] Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
[2] Water Engineering and Science Research Institute (WESRI), University of Tabriz, Tabriz, Iran

Abstract

Accurate forecasting of reference evapotranspiration (ET_o) is crucial for sustainable water resource management and precision agriculture. The present study evaluates three ET_o prediction methods: Random Forest (RF), Cartesian Genetic Programming (CGP), and Convolutional Neural Network-Graphics Processing Unit (CNN-GPU) across time intervals of 1 to 364 days. Using dispersion analysis (scatter/violin plots) and accuracy metrics (RMSE, MAE, R^2, SI), it was seen that the RF and CNN-GPU models consistently outperform CGP, particularly at extended horizons. At 364 days, CNN-GPU achieved the highest accuracy (RMSE: 0.678 mm/day, R^2: 0.874), while RF maintained robust performance (RMSE: 0.683 mm/day, R^2: 0.872) and minimal dispersion (SI: 0.244–0.278). In contrast, CGP exhibited slightly higher error indices (RMSE: 0.702mm/day) and greater variability. Statistical validation via t-tests, F-tests, and ANOVA revealed significant differences in performance, especially at longer lags ($p < 0.05$), with CNN-GPU often showing superior accuracy. Time-series analyses further confirmed that RF and CNN-GPU effectively capture seasonal ET_o trends, while CGP struggles with increased lag.

Keywords: ET_o prediction, random forest, CNN-GPU, genetic programming, dispersion analysis, time intervals.

1 Introduction

Reference evapotranspiration (ET_o), an important parameter in agricultural water management, represents the evaporative demand of the atmosphere independent of crop type, growth stage, or management practices. Allen et al. [2] argued that the Penman-Monteith equation adopted by FAO can serve as a reference standard formula for ET_o estimation under wide climatic conditions. This model closely aligns with spatial and temporal climatic patterns in regions with varying elevations and climates, emphasizing its adaptability [22]. Accurate ET_o prediction is critical for irrigation planning, water resource allocation, and sustainable agricultural practices, especially in regions with limited water resources [6, 23]. Given the impact of climate change

Submitted: 06 May 2025
Accepted: 25 August 2025
Published: 15 September 2025

Vol. 2, **No.** 3, 2025.
10.62762/TETAI.2025.125348

***Corresponding author:**
✉ Mostafa Sadeghzadeh
sadeqzadeh.m@gmail.com

Citation
Sadeghzadeh, M., Shiri, J., & Karimi, S. (2025). Performance Evaluation of ETo Prediction Methods: Dispersion Analysis and Accuracy Criteria Across Time Intervals. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 2(3), 182–191.

on precipitation patterns and increasing temperatures, accurate ET_o prediction is becoming increasingly important to mitigate drought impacts and ensure food security [21].

Traditional methods such as the Penman-Monteith equation require extensive meteorological data that are not always available. So, various models relying on fewer input variables have been developed for ET_o calculation. In recent years, machine learning (ML)-based techniques have gained popularity for hydrological parameters simulation due to their ability to model complex nonlinear phenomena without requiring explicit knowledge of the governing physical processes [13, 16]. Among these methods, genetic programming (GP) and random forest have shown promising results and have been widely applied in this context [11, 24]. Further, deep learning techniques, especially convolutional networks (CNNs), have been employed for predicting ET_o with limited input data [7, 8].

Genetic programming evolves computer programs to perform user-defined tasks [14]. In the context of ET_o prediction, GP has been used to develop empirical models that relate meteorological variables to ET_0 [13, 24]. Different variants of GP, such as grammatical evolution (GE), Cartesian genetic programming (CGP), and multi-objective genetic programming (MOGP), have been investigated to improve prediction accuracy and model interpretability [9, 12]. For example, gene expression programming (GEP) has been successful in predicting ET0 under different climate conditions due to its ability to produce simple and accurate models [25].

Random forest (RF) is an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting [4]. The ability of RF to handle large datasets and capture complex interactions makes it particularly suitable for time series forecasting tasks [3, 10].

Despite significant progress in ML-based models for ET_o prediction, comprehensive comparisons across diverse forecasting horizons remain limited, particularly for long-term predictions. Existing research has predominantly focused on short-term forecasts, often overlooking the accuracy and reliability of models over extended timescales. Moreover, the performance of GP variants, such as Cartesian GP (CGP), has not been thoroughly benchmarked against established ML-based techniques. This study addresses these gaps by evaluating three ET_o prediction methods, namely, RF, CGP, and CNN-GPU across timescales ranging from 1 to 364 days. Through dispersion analysis (scatter and violin plots), accuracy metrics (RMSE, MAE, R^2, SI), statistical validation (t-tests, F-tests, ANOVA), and time-series analysis, a detailed assessment of the models' performances were provided, with a particular emphasis on long-term forecasting capabilities.

2 Material and methods

2.1 Climatic characteristics and ET_o patterns of the study area

Data from two counties of Meshkinshahr and Ahar, located in Ardabil and East Azerbaijan provinces, Iran were utilized to evaluate the proposed methodologies. This region has a semi-arid and mountainous climate, with an average annual rainfall of about 300 to 400 mm and an average annual temperature of $12°C$ (see Figure 1). The ET_o series shown in Figure 1 represents the average ET_o calculated for the two study areas, Meshkinshahr and Ahar. This series reflects the mean ET_o values of these two locations and is plotted to provide a general overview of evapotranspiration patterns in the study region.

Daily ET_o data were collected from a meteorological station for the period 1 January 2015 to 20 March 2023. The data included ET_o values calculated using the standard FAO Penman-Monteith method, derived from meteorological variables e.g. air temperature, relative humidity, wind speed and solar radiation [24].

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

In this formula, ET_o stands for the reference evapotranspiration (mmd^{-1}) and R_n is the net solar radiation at the vegetative surface (MJ $\text{m}^{-2}\text{d}^{-1}$). G is the soil heat flux (MJ $\text{m}^{-2}\text{d}^{-1}$), T denotes the mean air temperature ($°C$), u_2 shows the wind speed at two meters above the ground (m/s), e_a and e_s represent the actual and saturated vapor pressures, respectively (kPa), Δ is the slope of the vapor pressure curve (kPa/$°C$), and γ shows the psychrometric constant (kPa/$°C$).

2.2 Overview of CGP, RF, and CNN-GPU

2.2.1 *Cartesian Genetic Programming (CGP)*

Cartesian Genetic Programming (CGP) employs a straightforward integer-based genetic representation

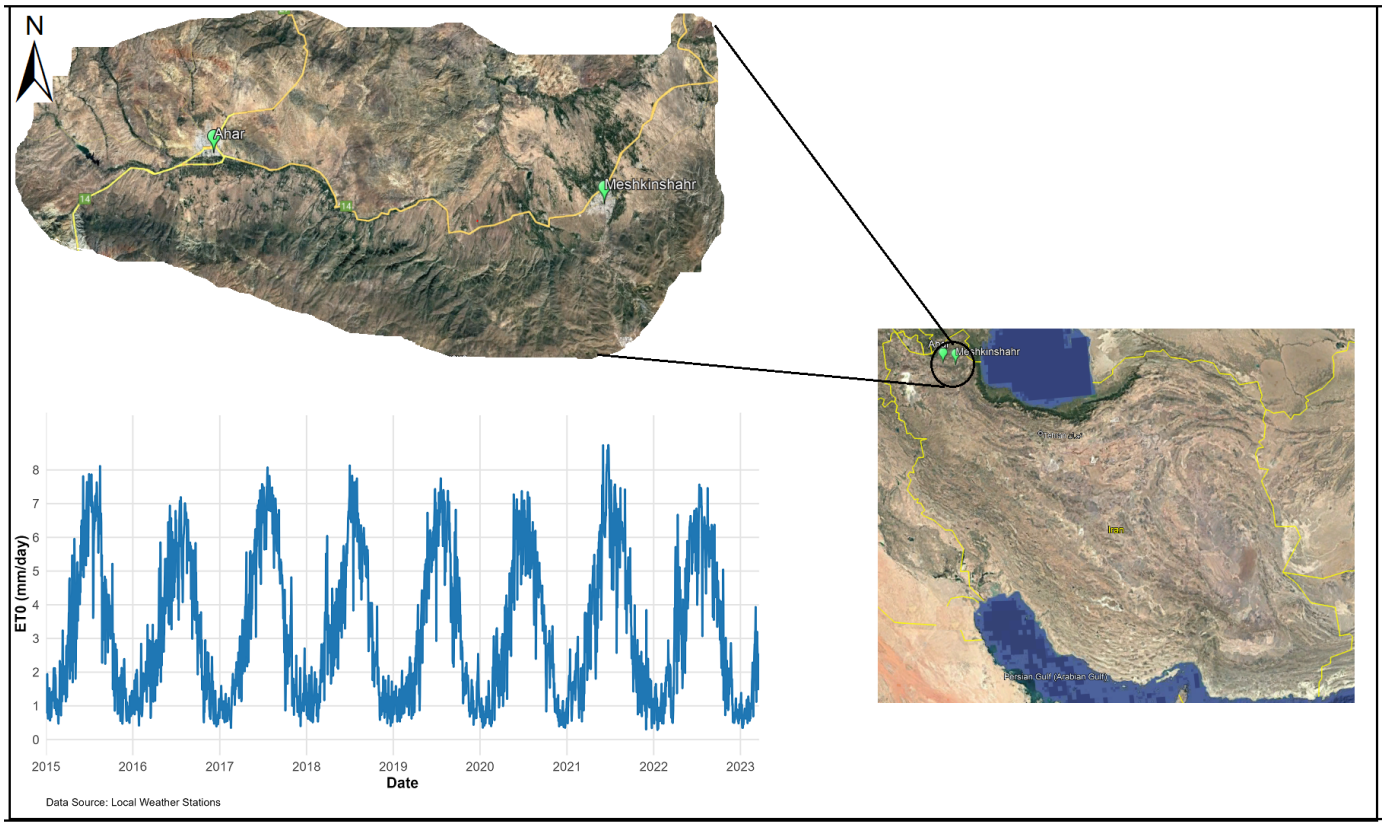

Figure 1. Study area and average ET_o time series.

of programs, structured as directed graphs [18]. These graph-based representations are highly versatile, supporting both cyclic and acyclic structures, and are applicable to diverse domains. Numerous studies have demonstrated that CGP outperforms other GP techniques in efficiency, while its simplicity facilitates its implementation [19, 20].

2.2.2 *Random Forest (RF)*

Random Forest (RF), a powerful ensemble learning approach, amplifies the predictive capability of multiple weak learners, such as single decision trees or perceptrons, through a collective decision-making process [1]. Introduced by Leo Breiman [4] in 2001, RF generates a series of unpruned classification or regression trees, each constructed from randomly sampled training data and feature subsets. The ensemble's predictions are synthesized via majority voting for classification tasks or averaging for regression tasks. RF demonstrates remarkable efficacy in datasets with a high number of variables relative to observations, making it a versatile tool for diverse, large-scale computational challenges. Its adaptability to specialized learning tasks and its capacity to evaluate variable significance further enhance its utility [4, 5]. By integrating numerous randomized decision trees and consolidating their outputs, RF delivers exceptional accuracy and robustness, establishing it as a cornerstone method in machine learning research.

2.2.3 *CNN-GPU model for ET_o prediction*

ET_o prediction was performed using CNN and accelerated by graphics processing units (GPU). The CNN model was designed to process time series data, especially meteorological [15] variables affecting ET_o. The architecture consisted of multiple one-dimensional convolutional layers with ReLU activation and max pooling layers to extract local features. The output was passed to fully connected layers to produce the final forecast. The model was trained with historical data and the mean square error was minimized with the Adam optimizer. Key equations involved in the convolution operation:

$$y_t = \text{ReLU}\left(\sum_{i=0}^{k-1} w_i x_{t+i} + b\right) \qquad (2)$$
$$Relu = max(0, x)$$

where y_t is output at time t after applying the ReLU activation function, ReLU is modified linear unit

activation function, k is Kernel size (number of filter weights), w_i is weight of the filter in position, x_{t+i} is input value at time $t+i$ and b is bias added to the weighted sum. The CNN model was implemented using the TensorFlow framework in Python with the Keras API, and the computations were accelerated on a GPU-accelerated system. The training process for the entire dataset took approximately 15 minutes at all time intervals. To quantify the performance gain, the same training process was run on a standard CPU. The CPU-based training took more than 60 minutes, indicating that GPU acceleration provides a performance speedup of over 5 times.

2.3 Evaluation Criteria

The performances of the applied models were evaluated using the following indices:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(ET_{o,i}^{\text{observed}} - ET_{o,i}^{\text{predicted}}\right)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(ETo_{predicted,i} - ETo_{observed,i}\right)^2} \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|ETo_{predicted,i} - ETo_{observed,i}\right| \quad (5)$$

$$\text{SI} = \frac{\text{RMSE}}{\overline{ET}_o} \quad (6)$$

where ET_o^{observed} and $ET_o^{\text{predicted}}$ are the observed and predicted values, respectively and $\overline{ET}_o$ is average ET_o value over the study period.

2.4 Model development

The data were split into training and test sets in a ratio of 80-20. For each forecast interval (lags 1, 11, 22, 91, 181, and 364 days), input matrices were constructed using ET_o values at times past the specified lag. Specifically, for a lag i, the input variables consisted of ET_o values at times $t-k$ where $k \leq i$. Each method was trained on the training set and evaluated on the test set. To ensure the robustness of the results, Monte Carlo simulations with 1000 iterations were performed to estimate 95% confidence intervals [17].

2.5 Statistical analyses

To examine statistical differences between methods, two-sample t-tests and F-tests were performed to compare means and variances of predictions in pairs [26]. Also, analysis of variance (ANOVA) was performed to assess overall differences between methods at each lag. These tests were performed at a significance level of 0.05 to identify significant differences. Figure 2 presents the schematic flowchart of the present research.

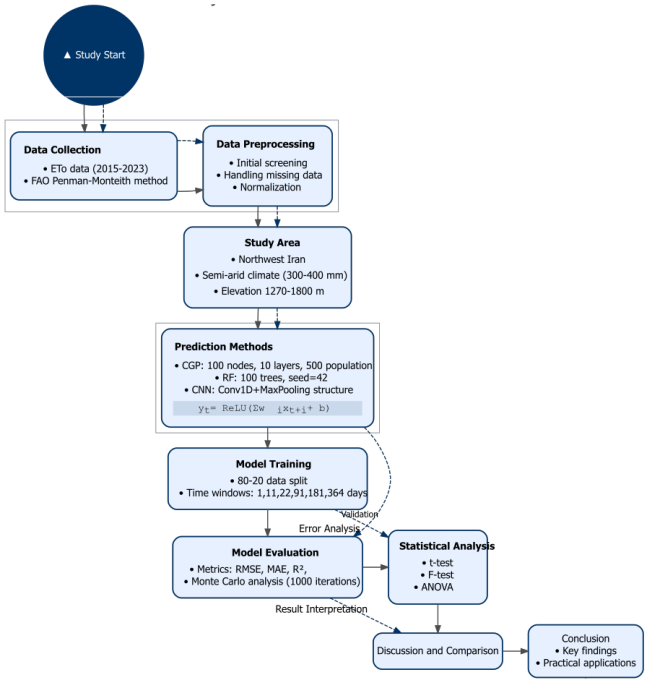

Figure 2. Schematic diagram of research stages.

3 Results and discussion

3.1 Performance evaluation of ET_o prediction methods: dispersion analysis and accuracy criteria in different time intervals

The evaluation results are presented in Figure 3 and Table 1. Figure 3 shows scatter plots of predicted and target ET_o values for the adopted time intervals (1, 11, 22, 91, 181, and 364 days). To provide context for the model configurations, a network with 100 nodes and up to 10 levels of recurrence, with a population of 500 and 50 generations was used for the CGP model. Further, RF regressor in the scikit-learn library was used with 100 trees (n_estimators=100) and a random seed of 42 for repeatability.

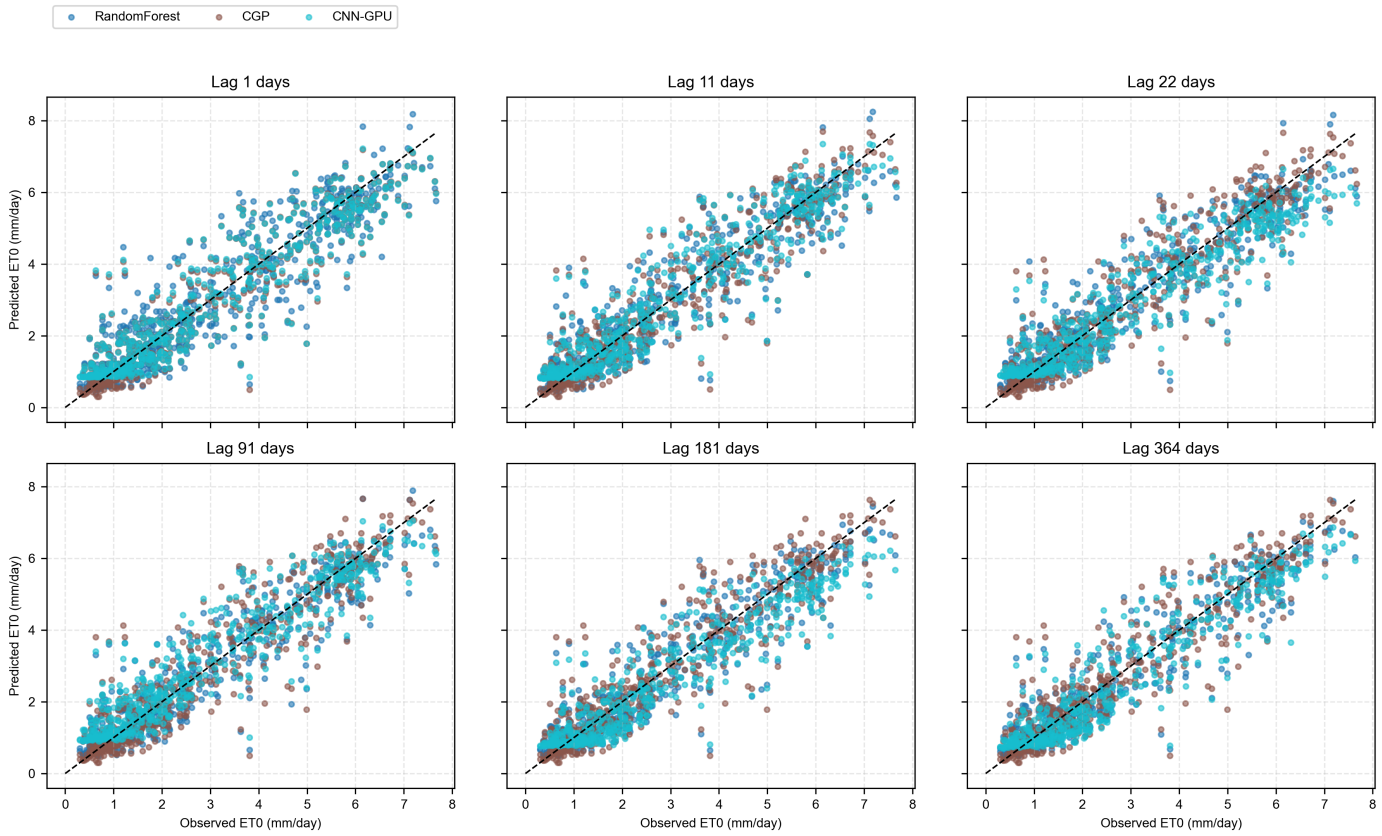

Figure 3. Scatter plots of ET_o forecasts over different time periods.

In Figure 3, each chart is plotted with a 1:1 reference line (solid line); points positioned nearer to this line represent more accurate predictions. At shorter lags (1, 11, and 22 days), the points for all methods are generally close to the 1:1 line with minimal scatter. CNN-GPU (turquoise) and RF show slightly tighter clustering compared to CGP at these lags, with CNN-GPU appearing to have the least scatter at lag 1 day. As the lag increases to 91, 181, and 364 days, the scatter of points increases slightly for all methods, but they remain relatively close to the 1:1 line. At longer lags, RF and CNN-GPU continue to show less deviation compared to CGP, which exhibits slightly more scatter, especially at 364 days interval.

Table 1 summarizes the evaluation metrics (RMSE, MAE, R^2, and SI) for each method across different prediction intervals. For the 1-day prediction, CNN-GPU achieves the best performance with an RMSE of 0.706 and R^2 of 0.872, whereas RF yields a higher RMSE of 0.803 and R^2 of 0.835. At 11 days, CNN-GPU again outperforms other methods with an RMSE of 0.694 and R^2 of 0.876, while RF records an RMSE of 0.735. For the 22-day horizon, RF and CNN-GPU produce nearly identical results, with RMSE values of 0.711 and 0.710, respectively. In contrast, at 91 days, RF delivers the most accurate predictions, achieving an RMSE of 0.687 and R^2 of 0.875, compared to CGP with an RMSE of 0.701. At 181 days, RF continues to lead with an RMSE of 0.688 and R^2 of 0.868, while CNN-GPU shows a slightly higher RMSE of 0.707. Finally, for the 364-day interval, CNN-GPU provides the best performance with an RMSE of 0.678 and R^2 of 0.874, whereas CGP records the highest RMSE of 0.702. The SI metric further indicates that all methods exhibit low relative dispersion, with CNN-GPU achieving the lowest SI at shorter horizons (e.g., 0.241 at 11 days) and RF maintaining lower dispersion at longer horizons (e.g., 0.244 at 91 days).

Overall, RF and CNN-GPU perform consistently across all prediction intervals, with CNN-GPU showing a slight edge at shorter lags (1 and 11 days) and RF performing better at intermediate lags (91 and 181 days). At the longest intervals (364 days), CNN-GPU again outperforms the others. CGP performs competitively but shows slightly more scatter at longer prediction intervals (e.g., 364 days). These results suggest that both RF and CNN-GPU are reliable for ET_o prediction across various time intervals, with CNN-GPU being particularly effective at shorter and

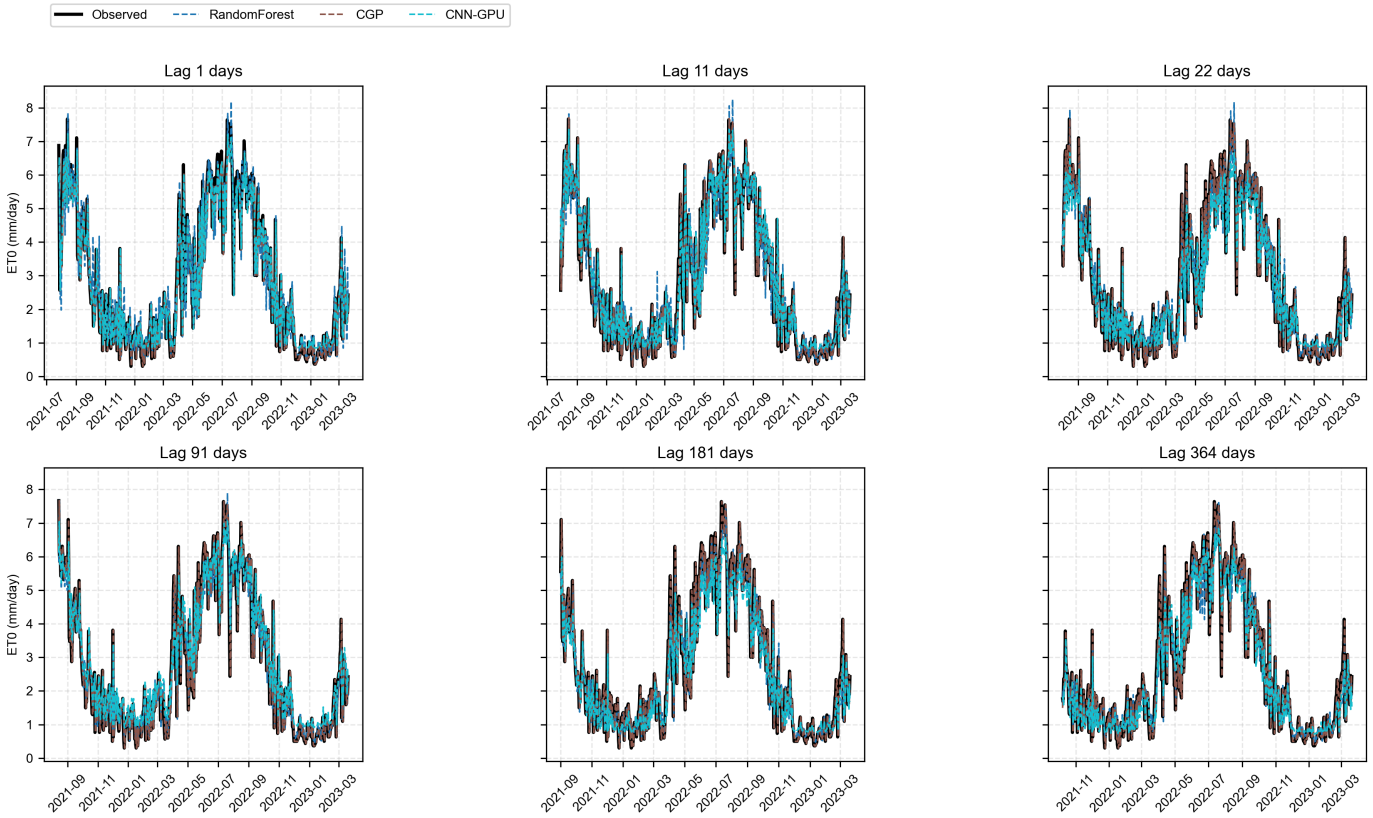

Figure 4. Comparing time series of ET_o forecasts with actual values.

very long lags.

3.2 Comparison of ET_o prediction time series for different methods at different lags

Figure 4 shows the predicted versus target ET_o values. At shorter intervals (1, 11, and 22 days), all methods closely follow the trend of the observed values, with minimal deviation at the extreme points. CNN-GPU and RF show slightly better alignment with the observed data compared to CGP, especially at lag 1 day. As the intervals increase to 91, 181, and 364 days, the predictions of all methods remain relatively close to the observed values, though slight deviations become more noticeable at the extreme points. RF and CNN-GPU continue to track the target data more closely than CGP, which shows slightly larger deviations at longer intervals, particularly at 364 days interval.

3.3 Statistical analysis of significant differences in the performance of ET_o prediction models using t and F tests

Figure 5 shows a bar graph of the statistical differences between pairs of ET_o prediction methods using t-tests and F-tests. The statistics are displayed on the left y-axis, and the p-values are shown in logarithmic

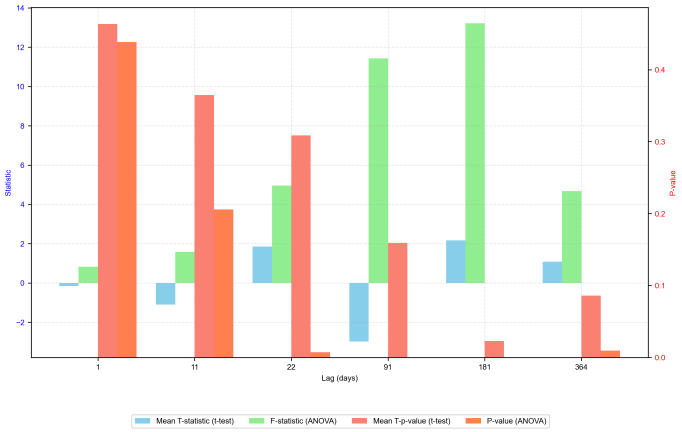

Figure 5. Results of t and F statistical tests for comparing prediction models.

form (-log10) on the right y-axis. The significance threshold line (p=0.05, equivalent to -log10(p)=1.3) can be inferred for interpretation.

3.4 Overall evaluation of statistical analysis for ET_o models

Statistical analysis using t-tests, F-tests, and ANOVA was conducted to assess significant differences in the performance of RF, CGP, and CNN-GPU across six prediction intervals (1, 11, 22, 91, 181, and

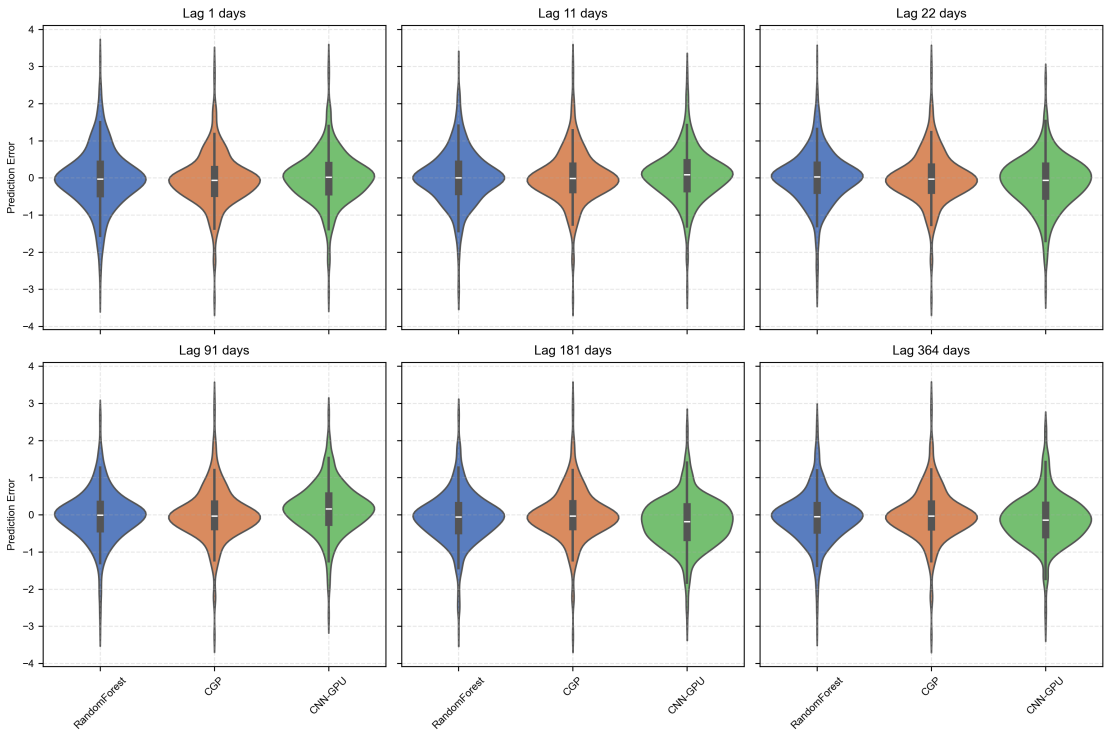

Figure 6. Violin diagrams of the distribution of forecast errors over time intervals.

Table 1. Performance evaluation criteria for RF, CGP, and CNN-GPU models.

Method	Lag	RMSE	MAE	R^2	SI
RF	1	0.803	0.577	0.835	0.278
CGP	1	0.708	0.494	0.872	0.245
CNN-GPU	1	0.706	0.502	0.872	0.244
RF	11	0.735	0.523	0.861	0.255
CGP	11	0.713	0.491	0.869	0.247
CNN-GPU	11	0.694	0.492	0.876	0.241
RF	22	0.711	0.502	0.870	0.247
CGP	22	0.704	0.486	0.873	0.244
CNN-GPU	22	0.710	0.525	0.871	0.246
RF	91	0.687	0.488	0.875	0.244
CGP	91	0.701	0.481	0.870	0.249
CNN-GPU	91	0.694	0.508	0.873	0.246
RF	181	0.688	0.489	0.868	0.253
CGP	181	0.704	0.482	0.862	0.259
CNN-GPU	181	0.707	0.535	0.861	0.260
RF	364	0.683	0.484	0.872	0.259
CGP	364	0.702	0.477	0.865	0.266
CNN-GPU	364	0.678	0.501	0.874	0.257

364 days). At shorter lags (1 and 11 days), t-tests and F-tests revealed no significant differences between the methods, with p-values consistently above 0.05 (e.g., ANOVA p-values of 0.438 and 0.206, respectively), indicating comparable performance and variance among the models. However, starting at 22 days, significant differences emerged, particularly involving CNN-GPU. For instance, at 22 days, t-tests showed significant differences for RF vs. CNN-GPU (p=0.0056) and CGP vs. CNN-GPU (p=0.0075), with ANOVA confirming overall significance (p=0.0071). This trend intensified at longer lags (91, 181, and 364 days), where CNN-GPU frequently outperformed others, as evidenced by highly significant t-test results (e.g., p=4.86E-07 for CGP vs. CNN-GPU at 181 days) and ANOVA p-values as low as 1.18E-05 (91 days) and 2.02E-06 (181 days). Notably, RF vs. CGP showed no significant differences across most lags (e.g., p=0.477 at 91 days), suggesting closer performance between these two methods.

The F-tests consistently indicated no significant differences in variance across all lags (p-values > 0.05, F-statistics near 1), implying similar dispersion in predictions among the methods. The ANOVA results highlighted that significant performance differences among the three methods became more pronounced at longer lags (22 days and beyond), representing the impact of method selection on prediction accuracy for extended forecasts. Overall, CNN-GPU demonstrated a consistent edge at shorter (1 and 11 days) and very

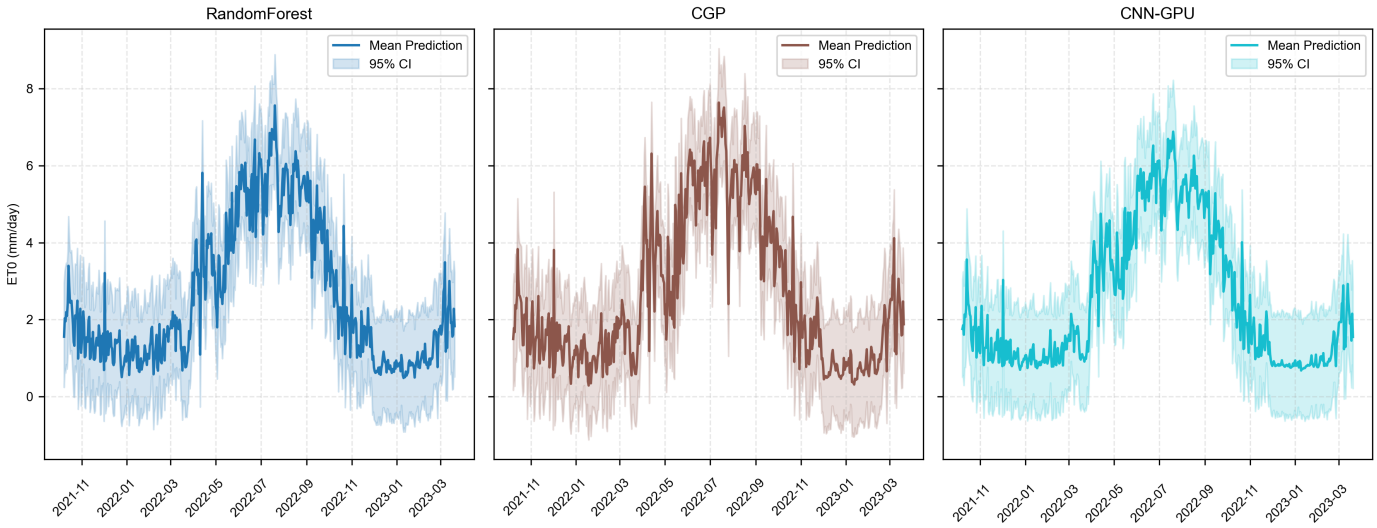

Figure 7. Average ET_o forecasts and 95% confidence interval over a 364-day period.

long lags (364 days), while RF performed robustly at intermediate lags (91 and 181 days). These findings suggest that CNN-GPU and RF are reliable choices for ET_o prediction, with CNN-GPU being particularly effective for short-term and long-term forecasts.

3.5 Analysis of the distribution of forecast errors at different time intervals

Figure 6 presents violin plots illustrating the distribution of forecast errors for the applied models. The width of each violin reflects the error density, while the central box plot indicates the median and interquartile range (IQR).

RF demonstrated the most stable performance across all intervals, with the smallest error ranges and distributions consistently centered around zero, making it the most reliable for predicting ET_o. CGP exhibited the greatest instability, with error dispersion escalating at longer lags, reaching a range of ± 3.31 at Lag 364. CNN-GPU showed intermediate performance, with wider errors than RF but less variability than CGP. These findings highlight Random Forest's consistent stability for ET_o forecasting across various time horizons, particularly at intermediate lags, and CNN-GPU's strong performance at short and very long lags, while CGP exhibits increasing inaccuracy at extended timescales.

3.6 Performance analysis of ET_o prediction methods at 364 delays

Figure 7 displays the mean forecasts and 95% confidence intervals of the models for the period from October 2021 to March 2023. All methods exhibit similar seasonal patterns with a peak in mid-2022. CGP has the highest mean forecast (maximum 7.64 mm/day), followed closely by RF (maximum 7.56 mm/day), while CNN-GPU has a lower maximum mean forecast (6.88 mm/day). However, RF demonstrates greater precision with a confidence interval width of 2.68 mm/day, compared to CGP's wider interval of 2.74 mm/day, though CNN-GPU has the narrowest interval at 2.61 mm/day. All methods produce unrealistic negative values at their lower bounds, with CGP showing the most extreme negative value (-1.12 mm/day), followed by RF (-0.92 mm/day), and CNN-GPU (-0.57 mm/day). Despite CGP's higher maximum forecast, RF strikes a better balance between accuracy and lower uncertainty, while CNN-GPU provides the tightest confidence intervals, indicating the least variability in predictions. Overall, RF offers a strong performance with high accuracy and reasonable consistency, though all methods struggle with physical consistency due to negative lower bounds.

4 Conclusion

This study provides a comprehensive evaluation of ET_o forecasting methods, demonstrating the outstanding performance of both RF and CNN-GPU across a range of timescales, from short to long. While GP-based models, such as CGP, show competitive accuracy over shorter intervals, their errors and dispersion increase significantly with longer forecasting horizons, as evidenced by wider error distributions. RF maintains exceptional stability with the tightest error distributions across all timescales, whereas CNN-GPU achieves superior accuracy at specific intervals, notably the 364-day timescale with the lowest RMSE (0.678).

Statistical analyses (t-tests, F-tests, ANOVA) confirm significant performance differences, particularly at longer timescales ($p < 0.05$), underscoring the critical importance of method selection for extended forecasts. RF's ability to model complex nonlinear relationships and seasonal patterns, supported by narrow confidence intervals and strong alignment with observations, highlights its reliability for practical applications. However, all methods, including RF and CNN-GPU, exhibit a standard limitation by occasionally producing spurious negative values at the prediction lower bound, indicating challenges in ensuring physical consistency. CGP, while promising for short-term use, proves inadequate for long-term dependencies, generating unrealistic outputs, such as negative ET_o values, alongside wider error distributions and high computational complexity, which limits its utility in large-scale water management. This physical inconsistency could be addressed through post-processing (setting negatives to zero) or by integrating constraints directly into model architectures (e.g., ReLU/Softplus activation functions). Future research should enhance predictive accuracy and reliability by exploring hybrid models like CNN-LSTM to capture spatiotemporal dependencies and implementing advanced uncertainty quantification techniques (e.g., Bayesian deep learning, quantile regression) to provide robust confidence intervals for improved risk assessment in hydrology and precision agriculture. Such advancements will solidify machine learning's role in addressing critical environmental challenges.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and buildings, 147*, 77-89. [Crossref]

[2] Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome, 300*(9), D05109.

[3] Ao, Y., Li, H., Zhu, L., Ali, S., & Yang, Z. (2019). The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. *Journal of Petroleum Science and Engineering, 174*, 776-789. [Crossref]

[4] Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32. [Crossref]

[5] Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In *Ensemble machine learning* (pp. 157-175). Springer, New York, NY. [Crossref]

[6] Djaman, K., Balde, A. B., Sow, A., Muller, B., Irmak, S., N'Diaye, M. K., ... & Saito, K. (2015). Evaluation of sixteen reference evapotranspiration methods under sahelian conditions in the Senegal River Valley. *Journal of Hydrology: regional studies, 3*, 139-159. [Crossref]

[7] e Lucas, P. D. O., Alves, M. A., e Silva, P. C. D. L., & Guimaraes, F. G. (2020). Reference evapotranspiration time series forecasting with ensemble of convolutional neural networks. *Computers and electronics in agriculture, 177*, 105700. [Crossref]

[8] Ferreira, L. B., & da Cunha, F. F. (2020). Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Computers and electronics in agriculture, 178*, 105728. [Crossref]

[9] Garcia-Garcia, C., Morales-Reyes, A., & Escalante, H. J. (2023). Continuous Cartesian Genetic Programming based representation for multi-objective neural architecture search. *Applied Soft Computing, 147*, 110788. [Crossref]

[10] Huang, H., Pouls, M., Meyer, A., & Pauly, M. (2020, September). Travel time prediction using tree-based ensembles. In *International Conference on Computational Logistics* (pp. 412-427). Cham: Springer International Publishing. [Crossref]

[11] Karimi, S., Sadraddini, A. A., Nazemi, A. H., Xu, T., & Fard, A. F. (2018). Generalizability of gene expression programming and random forest methodologies in estimating cropland and grassland leaf area index. *Computers and Electronics in Agriculture, 144*, 232-240. [Crossref]

[12] Khayyam, H., Jamali, A., Assimi, H., & Jazar, R. N. (2019). Genetic programming approaches in design and optimization of mechanical engineering applications. In *Nonlinear Approaches in Engineering Applications: Automotive Applications of Engineering Problems* (pp. 367-402). Cham: Springer International Publishing. [Crossref]

[13] Kisi, O., & Guven, A. (2010). Evapotranspiration modeling using linear genetic programming technique. *Journal of Irrigation and Drainage Engineering,*

136(10), 715-723. [Crossref]

[14] Koza, J. R. G. P. (1992). On the programming of computers by means of natural selection. *Genetic programming*.

[15] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444. [Crossref]

[16] Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2017). Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and electronics in agriculture, 139*, 103-114. [Crossref]

[17] Metropolis, N., & Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association, 44*(247), 335-341.

[18] Miller, J. F. (2020). Cartesian genetic programming: its status and future. *Genetic Programming and Evolvable Machines, 21*(1), 129-168. [Crossref]

[19] Miller, J. F., & Harding, S. L. (2009, July). Cartesian genetic programming. In *Proceedings of the 11th annual conference companion on genetic and evolutionary computation conference: late breaking papers* (pp. 3489-3512). [Crossref]

[20] Miller, J., & Turner, A. (2015, July). Cartesian genetic programming. In *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation* (pp. 179-198). [Crossref]

[21] Pereira, L. S., Allen, R. G., Smith, M., & Raes, D. (2015). Crop evapotranspiration estimation with FAO56: Past and future. *Agricultural water management, 147*, 4-20. [Crossref]

[22] Sadeghzadeh, M., Shiri, J., Karimi, S., & Majnooni, A. (2024). Analysis of factors affecting evapotranspiration zoning. *Environmental Science and Pollution Research, 31*(29), 42295-42313. [Crossref]

[23] Sentelhas, P. C., Gillespie, T. J., & Santos, E. A. (2010). Evaluation of FAO Penman–Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. *Agricultural water management, 97*(5), 635-644. [Crossref]

[24] Shiri, J. (2019). Modeling reference evapotranspiration in island environments: Assessing the practical implications. *Journal of Hydrology, 570*, 265-280. [Crossref]

[25] Shiri, J., Kişi, Ö., Landeras, G., López, J. J., Nazemi, A. H., & Stuyt, L. C. (2012). Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). *Journal of Hydrology, 414*, 302-316. [Crossref]

[26] Snedecor, G. W., & Cochran, W. G. (1989). Arc sine transformation for proportions. Statistical Methods.