

# Advanced evapotranspiration forecasting in Central Italy: Stacked MLP-RF algorithm and correlated Nystrom views with feature selection strategies

Francesco Granata<sup>\*</sup>, Fabio Di Nunno, Giovanni de Marinis

*University of Cassino and Southern Lazio, Department of Civil and Mechanical Engineering (DICEM), Via Di Biasio, 43, 03043 Cassino, Frosinone, Italy*



## ARTICLE INFO

**Keywords:**  
 Reference Evapotranspiration  
 Forecasting  
 Machine Learning  
 Feature Selection  
 Stacked Model  
 Correlated Nystrom Views

## ABSTRACT

Evapotranspiration is a key parameter in hydrology, particularly in the field of water resources management. Reference evapotranspiration (ET<sub>0</sub>) stands as a crucial metric, embodying the influence of climate on water loss from both soil and plant surfaces. The accurate forecasting of future ET<sub>0</sub> values is paramount for informed decision-making in agricultural practices and water supply planning. The anticipation of evapotranspiration variations supports optimized irrigation, drought assessment, and efficient water allocation. Employing innovative algorithms, specifically the Multilayer Perceptron-Random Forest (MLP-RF) Stacked Model and the Correlated Nystrom Views (XNV), this study focuses on predicting ET<sub>0</sub> up to 60 days ahead in Agro Pontino, an area in Central Italy known for its flourishing agricultural production in the Mediterranean Europe. A Radial Basis Function (RBF) Neural Network serves as a benchmark. Taking into account firstly all available weather-climatic variables and subsequently adopting two different strategies for reducing the variables, based respectively on Principal Component Analysis and feature selection using Particle Swarm Optimization, three distinct sets of input variables were considered. The models based on the complete set of exogenous climatic variables demonstrated superior accuracy. However, even models relying on only mean temperature, maximum relative humidity, and shortwave solar radiation as inputs produced good results, also for the 60-day forecasting horizon, with Kling-Gupta Efficiency (KGE) and Mean Absolute Percentage Error (MAPE) equal to 0.98 and 8.356 %, respectively, in the case of MLP-RF Stacked Model. The latter consistently outperformed XNV and RBF across various combinations of input variables and forecasting horizons. Notably, the reduction in accuracy with extended forecasting horizons was mild, suggesting the potential for accurate results over significantly more extended horizons. These forecasting models facilitate precise irrigation scheduling, minimizing water wastage and conserving resources. This targeted approach enhances crop yields, quality, and environmental sustainability, rendering agriculture economically viable and adaptable to climate variability.

## 1. Introduction

Reference Evapotranspiration (ET<sub>0</sub>) is a preeminent parameter in agricultural and environmental sciences, embodying a fundamental metric for appraising atmospheric water demand. Encompassing the amalgamated influences of evaporation from a standardized reference surface and transpiration from an idealized crop, ET<sub>0</sub> is pivotal in unravelling the intricacies of water availability and consumption across diverse ecosystems. Accurately forecasting reference evapotranspiration is imperative for shaping sustainable water management strategies, optimizing irrigation practices, and advancing our understanding of hydrological processes on a broader scale (Gocic et al. 2015).

Reference evapotranspiration incorporates a set of meteorological

and physical parameters that collectively influence water loss from a given surface. Variables such as temperature, humidity, wind speed, and solar radiation intricately contribute to the complex equilibrium governing the transition of liquid water into vapour (Yu et al., 2020). A profound understanding of these meteorological factors not only facilitates but is also instrumental in the development of robust models for accurately forecasting ET<sub>0</sub> (Chia et al., 2020). The seamless integration of such models into scientific inquiries enables the precise quantification of water requirements for diverse crops, thereby assisting farmers and land managers in making judicious decisions regarding irrigation scheduling and resource allocation.

Beyond agriculture, accurate forecasting of reference evapotranspiration plays a pivotal role in evaluating the overall water balance within

\* Corresponding author.

E-mail addresses: [f.granata@unicas.it](mailto:f.granata@unicas.it) (F. Granata), [fabio.dinunno@unicas.it](mailto:fabio.dinunno@unicas.it) (F. Di Nunno), [demarinis@unicas.it](mailto:demarinis@unicas.it) (G. de Marinis).

ecosystems (Dou & Yang, 2018). This parameter serves as a foundational input in hydrological models, facilitating the assessment of watershed dynamics, groundwater recharge, and the overall resilience of ecosystems to changing climatic conditions. Researchers and policymakers rely on ETo predictions to make informed decisions about water allocation, land-use planning, and environmental conservation efforts.

The landscape of models employed to predict future evapotranspiration is characterized by distinct categories, each supported by different methodologies and theoretical frameworks. Classical empirical models, rooted in historical data and statistical relationships, form one category, relying on historical climatic patterns to project future evapotranspiration. Physically-based models constitute another category, employing complex mathematical representations of evapotranspiration's physical processes, often integrating meteorological variables such as temperature, humidity, wind speed, and solar radiation. Recent advancements have witnessed the emergence of hybrid models, fusing empirical and physically-based elements, leveraging the strengths of both approaches for enhanced predictive accuracy. Furthermore, machine learning (ML) models have gained prominence, harnessing computational algorithms to discern intricate patterns within vast datasets and adaptively refining predictions (Goyal et al., 2023). In recent years, they have found application in research related to the evaluation of ETo or actual evapotranspiration ( $ET_a$ ) models, as evidenced by several notable studies (Feng et al., 2017, Kisi & Alizamir, 2018, Xu et al., 2018, Dou & Yang, 2018, Granata, 2019, Malik et al., 2019, Tikhamarine et al., 2019, Zhao et al. 2019, Tikhamarine et al., 2020a, 2020b, Yamac & Todorovic 2020, Bai et al., 2021, Aghelpour & Norooz-Valashedi, 2022, Jia et al., 2023).

Until now, the application of ML and deep learning (DL) algorithms for short- and medium-term ETo forecasting has been comparatively infrequent. Trajkovic et al. (2003) implemented a sequentially adaptive Radial Basis Function (RBF) Neural Network to predict ETo in Nis, Serbia. Notably, their prediction model did not incorporate exogenous climate variables. In a similar vein, Landeras et al. (2009) conducted a comparative analysis between Artificial Neural Networks (ANN) and Autoregressive Integrated Moving Average (ARIMA) models for weekly average ETo prediction in the Alava region, Basque Country, Spain. In this research, climatic variables were also omitted, with the forecast model relying solely on lagged ETo values. Torres et al. (2011) utilized Multivariate Relevance Vector Machine (MRVM) and ANN for short-term ETo forecasting in Central Utah. Meanwhile, Luo et al. (2015) employed four ANNs for ETo prediction, exclusively incorporating predicted temperature data from public weather forecasts. Notably, both studies considered only temperature as an exogenous climatic input variable. Chen et al., (2020a) employed three deep learning models, Deep Neural Network (DNN), Temporal Convolutional Neural Network (TCN), and Long Short-Term Memory Neural Network (LSTM), to estimate daily ETo in the Northeast Plain of China, characterized by a warm temperate zone with continental and monsoon climate characteristics. Subsequently, the same authors expanded their work, developing a TCN incorporating Principal Component Analysis and Maximal Information Coefficient to predict crop evapotranspiration using a two-year dataset from lysimeters for maize under drip irrigation (Chen et al., 2020b). At the same time, Ferreira & da Cunha (2020a) harnessed the power of LSTM networks, one-dimensional convolutional neural networks (1D CNN), and a hybrid CNN-LSTM algorithm. This comparative study, conducted in Minas Gerais, a state in South-eastern Brazil characterized by a very warm climate, extended the forecasting horizon to multi-step predictions of daily ETo, spanning up to seven days. Granata & Di Nunno (2021) developed three predictive models based on Deep Learning, utilizing both NARX and LSTM networks, to forecast actual evapotranspiration in two distinct climatic zones, with prediction horizons ranging from 1 to 7 days. They discovered that LSTM models exhibited greater accuracy than NARX models in subtropical climatic conditions. Conversely, NARX models consistently yielded more precise results in semi-arid climates. The accuracy of the models was influenced by

various exogenous variables, with notable impacts from sensible heat flux and relative humidity. Additionally, the forecast horizon played a significant role, exerting distinct effects in different climatic conditions. Recently, Valipour et al. (2023) employed several hybrid models of deep learning and machine learning, specifically wavelet long short-term memory (WLSTM), wavelet group method of data handling (WGMDH), and wavelet genetic algorithm-adaptive neuro-fuzzy inference system (WGA-ANFIS), for forecasting reference evapotranspiration up to 10 days ahead at 30 sites across the United States. They achieved excellent results, with an average RMSE across the 30 sites for 1-day ahead forecasts equal to 0.541 mm/day.

The research conducted so far on the topic of forecasting future daily ETo has generally focused on short-term prediction horizons, seldom exceeding 7–15 days, consistently revealing a significant decline in accuracy with increasing forecast horizon. The primary objective of this study is to overcome this limitation by proposing highly accurate ETo prediction models with horizons extending up to 60 days. To achieve this goal, we aim to introduce two algorithms never before employed in ETo prediction: the Correlated Nystrom Views (XNV) and the MLP-RF Stacked Model. The performance of these algorithms will be compared with that of a model based on RBF, chosen as the benchmark algorithm. As ETo depends on numerous meteorological and climatic variables, two distinct variable reduction strategies are implemented: Principal Component Analysis (PCA) and feature selection using the Particle Swarm Optimization (PSO) algorithm. The impact of this predictor reduction on predictive accuracy is assessed. The chosen study area for this research is one of the most flourishing agricultural production regions in the Mediterranean area: the Agro Pontino (Fig. 1), located in Central-Southern Lazio, Italy. The decision to focus on one of the key sites in Southern Europe for the production of fruits and vegetables adds even greater practical relevance to the study.

## 2. Materials and methods

### 2.1. Study area and dataset

The Agro Pontino, located in the Lazio region, central Italy, is a distinctive agricultural area characterized by flat plains and reclaimed marshlands. Historically, this region was a malarial swamp until ambitious drainage projects in the 1930s transformed it into fertile agricultural land. Geographically, the Agro Pontino is marked by expansive plains that stretch between the Tyrrhenian Sea and the foothills of the Apennine Mountains. Rich, alluvial soils support a diverse range of crops, contributing to the area's agricultural significance. The climate is typically Mediterranean, characterized by hot, dry summers and mild, wet winters (Cioffi et al., 2017). This climate pattern, coupled with the fertile soils, makes the Agro Pontino conducive to a variety of agricultural activities. Agriculture in the Agro Pontino thrives with the cultivation of cereals, vegetables, and fruits. Olive orchards and vineyards dot the landscape, producing high-quality olive oil and wine. Livestock farming also contributes to the region's vibrant agricultural economy.

Consequently, a proper prediction of ETo in the Agro Pontino region is crucial due to its direct impact on water management and agricultural productivity. ETo influences the water needs of crops, and accurate predictions aid farmers in optimizing irrigation practices, ensuring efficient water use. Given the region's historical transformation from marshland to arable land, understanding evapotranspiration dynamics becomes paramount for sustainable agricultural practices. Effective management of this process contributes to the resilience of crops, particularly in the face of the Agro Pontino's Mediterranean climate, characterized by distinct seasons, further emphasizing the significance of precise evapotranspiration forecasts in promoting agricultural sustainability.

The current study relies on the MADIA (Meteorological variables for Agriculture: daily time series for the Italian Area, (Parisse et al., 2023) gridded dataset. This dataset provides the times series of the main agro-

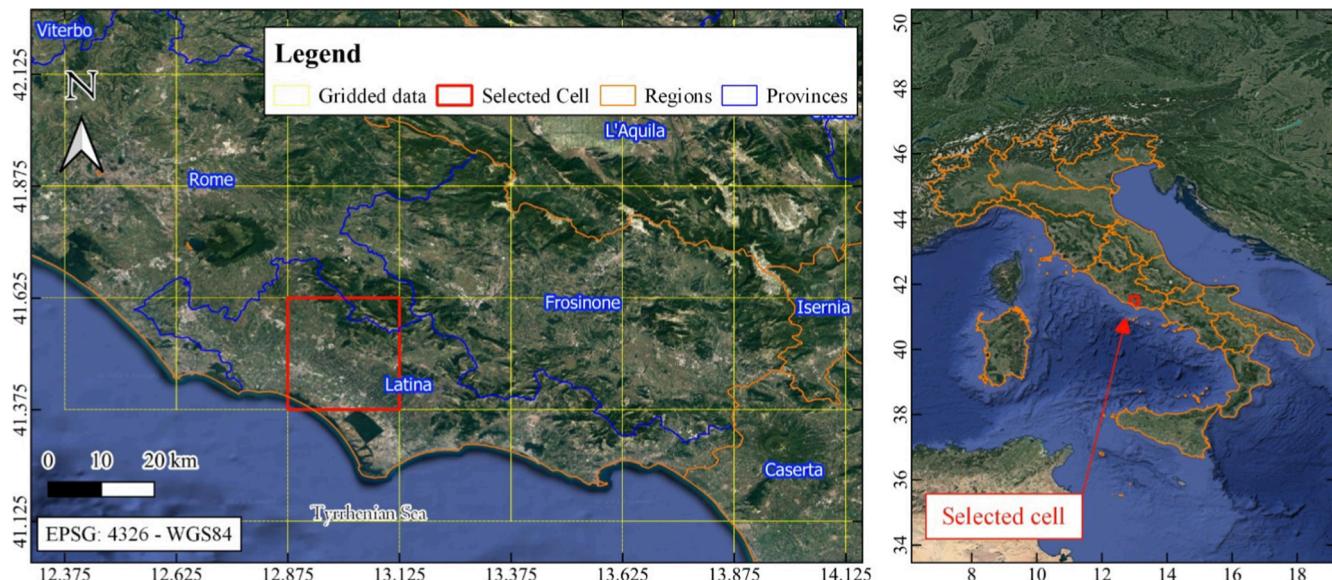


Fig. 1. Location of the selected cell in the Lazio territory.

meteorological variables derived from ERA5 hourly surface data, with a spatial resolution of 0.25 degrees, for the period 1981–2022. The dataset contains the daily time series of minimum ( $T_{\min}$ ), average ( $T_{\text{mean}}$ ) and maximum ( $T_{\max}$ ) air temperature, minimum ( $RH_{\min}$ ) and maximum ( $RH_{\max}$ ) air relative humidity, wind speed at the height of 10 m above the surface of the Earth ( $WS_{10}$ ), solar radiation downwards (shortwave radiation,  $Rs$ ), precipitation ( $P$ ) and  $ETo$  computed based on the FAO Penman-Monteith equation (Allen et al., 1998).

The selected cell, located near the Tyrrhenian Sea, encompasses the city of Latina and a vast area characterized by intensive cultivation. Fig. 2 presents the monthly mean values of meteorological variables for the selected cell. Air temperature,  $Rs$  and  $ETo$  exhibit a consistent pattern, showcasing higher values in summer and lower values in winter. Conversely,  $RH$ ,  $WS_{10}$ , and  $P$  demonstrate an opposite trend, featuring higher values in winter and lower values in summer. This seasonal variation underscores the dynamic interplay of climatic factors, with air temperature and solar radiation peaking during warmer months while relative humidity, wind speed, and precipitation reach their peak in the cooler winter period.

## 2.2. Principal Component analysis

PCA is a powerful statistical method widely employed for multivariate data analysis and dimensionality reduction. Originating from the seminal work of Pearson (1901) and subsequently refined by Hotelling (1933), PCA has become a cornerstone technique in various scientific disciplines.

At its core, PCA seeks to transform a high-dimensional dataset into a new coordinate system, capturing the maximum variance along its principal components. The first step involves the computation of the covariance matrix for the given dataset, expressing the relationships between variables. The subsequent stage entails the eigen-decomposition of the covariance matrix, yielding a set of eigenvectors and corresponding eigenvalues. These eigenvectors constitute the principal components, defining the directions in which the data exhibits the maximum variance. The eigenvalues quantify the amount of variance explained by each principal component, aiding in the selection of relevant components for dimensionality reduction.

Following data standardization to ensure comparability, the PCA

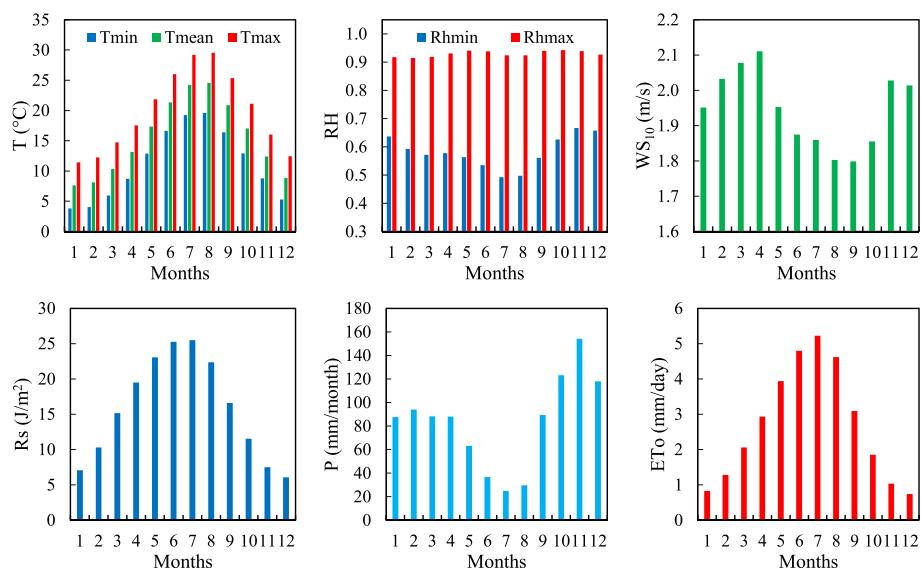


Fig. 2. Monthly mean values of the meteorological variables.

algorithm involves selecting the top  $k$  eigenvectors corresponding to the  $k$  largest eigenvalues. This selection forms the basis for a new subspace onto which the original data is projected, effectively reducing its dimensionality.

For the problem under consideration (Zhao et al., 2021), given all the input variables, namely:  $T_{\min}$ ,  $T_{\max}$ ,  $T_{\text{mean}}$ ,  $RH_{\min}$ ,  $RH_{\max}$ ,  $WS_{10}$ ,  $Rs$ ,  $P$ , the correlation matrix is shown in the following Table 1.

Selecting 5 principal components yields the highest eigenvalues, as reported in Table 2.

As evident from the table, the first 5 eigenvalues account for approximately 97 % of the total variance. The eigenvectors were subsequently derived, and from these, the new input variables were obtained, as presented in the following Table 3.

### 2.3. Feature selection by Particle Swarm optimization

Within the domain of regression problems, delineated by datasets  $D = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\}$  where  $X_i$  denotes feature vectors and  $y_i$  are the target values, the fundamental objective is to distil a subset of features  $S$  from the original set  $F = \{f_1, f_2, \dots, f_m\}$  that significantly contributes to the regression model's efficacy.

Inspired by collective behaviour in nature, PSO (Kennedy & Eberhart, 1995) operates as a population-based optimization algorithm. In the context of feature selection for regression, PSO unfolds through pivotal mechanisms:

- Particle Representation: Each particle in the swarm encapsulates a potential feature subset  $S_i = \{f_{i1}, f_{i2}, \dots, f_{ik}\}$  where  $k \leq m$ . The binary status of feature inclusion (1) or exclusion (0) is encoded in the particle's position  $X_i$ .
- Objective Function: The fitness or objective function  $f(X_i)$  evaluates the regression model's performance using the selected features, typically gauged through regression-specific metrics such as root mean squared error (RMSE).
- Velocity and Position Update: Particles iteratively refine positions based on historical best positions and the global best position. The velocity update equation is expressed as

$$v_{ij}^{(t+1)} = w \cdot v_{ij}^t + c_1 \cdot r_1 \cdot (p_{best-ij} - x_{ij}^t) + c_2 \cdot r_2 \cdot (g_{best-ij} - x_{ij}^t) \quad (1)$$

incorporating inertia weight  $w$ , acceleration coefficients  $c_1$  and  $c_2$ , and random values  $r_1$  and  $r_2$ .

The PSO algorithm iteratively hones feature subsets until a convergence criterion is met, be it a predetermined number of iterations or the satisfactory convergence of fitness values, signifying stability in the selected feature subset. The features selected here through the PSO are as follows:  $T_{\text{mean}}$ ,  $RH_{\max}$ ,  $Rs$ .

### 2.4. Radial basis function Neural network (RBF-NN)

The RBF-NN (Park & Sandberg, 1991) is structured as a three-layered feedforward network comprising an input layer, a hidden layer, and an output layer. The transformative power of the hidden layer is realized through radial basis functions, which project the input data into a

**Table 2**

First five eigenvalues of the correlation matrix.

Eigenvalue	Proportion	Cumulative
3.80099	0.47512	0.47512
1.64158	0.2052	0.68032
1.3647	0.17059	0.85091
0.52254	0.06532	0.91623
0.4338	0.05417	0.9704

**Table 3**

Attributes defined by Principal Component Analysis.

Attribute	Linear combination of the initial variables
1	-0.498T <sub>max</sub> -0.483T <sub>mean</sub> -0.451T <sub>min</sub> -0.443Rs + 0.275RH <sub>min</sub>
2	0.559P + 0.519RH <sub>min</sub> + 0.404RH <sub>max</sub> + 0.325T <sub>min</sub> + 0.237T <sub>mean</sub>
3	0.696WS <sub>10</sub> -0.611RH <sub>max</sub> + 0.294P-0.22RH <sub>min</sub> + 0.07 T <sub>min</sub>
4	-0.625RH <sub>max</sub> -0.481WS <sub>10</sub> -0.41Rs + 0.36 RH <sub>min</sub> + 0.218T <sub>min</sub>
5	0.751P-0.465WS <sub>10</sub> -0.414RH <sub>min</sub> + 0.158Rs-0.126T <sub>min</sub>

higher-dimensional space, thereby facilitating the regression task.

The radial basis function, a mathematical function centered at a specific point, exhibits an exponential decay as the distance from the center increases. The initialization of the Gaussian radial basis functions involves utilising the K-Means algorithm to identify optimal centers. The Gaussian function stands out as the predominant radial basis function, serving as the activation function for the hidden layer. The output layer of the RBF neural network is typically characterized by a linear combination of the activations from the hidden layer, with weights determined through either a least-squares approach or gradient descent.

Notably, the RBF-NN distinguishes itself by necessitating a reduced volume of training data compared to other neural networks. This advantage arises from the feature extraction capabilities of the hidden layer, which effectively reduces the dimensionality of the input data. Furthermore, the RBF-NN exhibits a diminished susceptibility to overfitting, a phenomenon where a model becomes excessively intricate, fitting the training data meticulously but faltering in generalization to new data.

### 2.5. Correlated Nystrom views (XNV)

The XNV algorithm (McWilliams et al., 2013) is a rapid semi-supervised solution for regression and classification tasks, effectively addressing the challenges posed by extensive datasets and the resource-intensive nature of acquiring labelled data. XNV is grounded in two pivotal concepts: the generation of two distinct views featuring computationally efficient random features and the incorporation of Canonical Correlation Analysis (CCA) for multiview regression on unlabeled data to bias regression towards valuable features.

Motivated by the computational costs inherent in handling large datasets, XNV offers an appealing alternative to kernel methods by harnessing the power of randomization. It delivers comparable generalization performance while substantially reducing computational costs, strategically targeting the intricate task of obtaining labelled data for expansive datasets.

**Table 1**  
Correlation matrix.

	T <sub>min</sub>	T <sub>max</sub>	T <sub>mean</sub>	RH <sub>min</sub>	RH <sub>max</sub>	WS <sub>10</sub>	Rs	P
T <sub>min</sub>	1.00	0.94	0.98	-0.17	0.05	-0.01	0.61	-0.03
T <sub>max</sub>	0.94	1.00	0.99	-0.39	0.02	-0.14	0.76	-0.21
T <sub>mean</sub>	0.98	0.99	1.00	-0.29	0.04	-0.08	0.70	-0.12
RH <sub>min</sub>	-0.17	-0.39	-0.29	1.00	0.42	0.08	-0.59	0.45
RH <sub>max</sub>	0.05	0.02	0.04	0.42	1.00	-0.25	-0.02	0.12
WS <sub>10</sub>	-0.01	-0.14	-0.08	0.08	-0.25	1.00	-0.17	0.41
Rs	0.61	0.76	0.70	-0.59	-0.02	-0.17	1.00	-0.37
P	-0.03	-0.21	-0.12	0.45	0.12	0.41	-0.37	1.00

The algorithm unfolds in a systematic two-step manner. Initially, it meticulously constructs two views using random features, applying the Nystrom method. Subsequently, Canonical Correlation Analysis is employed to bias the optimization procedure, with the aim of minimizing variance without a significant increase in bias, particularly effective when both views contain accurate estimators.

At its essence, the Nystrom method is designed to approximate kernel matrices, foundational to kernel methods, through an adept low-rank factorization strategy. Given the cubic computational complexity associated with exact kernel computations on extensive datasets, the efficiency of the Nystrom method becomes increasingly indispensable.

The method begins by selecting a subset, referred to as the “landmark set”, from the entire dataset. The kernel matrix is then computed solely for these selected landmarks, resulting in a substantial reduction in computational burden. The subsequent low-rank approximation is achieved by considering the cross-covariance between the entire dataset and the landmark set. Then, CCA plays a key role in revealing latent relationships between paired sets of variables and enhancing predictive modelling when confronted with data characterized by multiple perspectives or “views”.

The basic principle of CCA involves identifying linear combinations of variables within each view, termed canonical variables, maximizing the correlation between the views. In the field of multiview regression, where disparate yet potentially complementary information is available, CCA becomes instrumental in integrating these views for enhanced predictive performance.

CCA optimization aims to maximize the correlation between canonical variables, effectively capturing shared information between views. This emphasis on correlation aligns with the intuition that, in multiview scenarios, variables across views may exhibit interconnected patterns, and exploiting such correlations enhances the predictive utility of the model.

## 2.6. Multilayer Perceptron-Random Forest (MLP-RF) Stacked model

The Stacked Model integrates MLP, RF, and Elastic Net (EN) as its constituent model elements. The initial stage involves training the base models, MLP and RF, on the same dataset. Subsequently, the *meta*-learner EN utilizes the predictions generated by these base models as input features, steering the ensemble model toward its final output.

MLP (Murtagh, 1991), a prominent artificial neural network in supervised learning, features a layered architecture comprised of interconnected nodes or neurons. These layers process input data through a hierarchical series of transformations, culminating in the generation of predictions. The structural framework of MLP follows a specific pattern: the input layer receives raw data and transmits it to the initial hidden layer, where a sequence of nonlinear transformations occurs. Output from each hidden layer is successively forwarded to subsequent layers until the final prediction emerges from the output layer. Within each layer, adjustable weights and biases undergo refinement during training to optimize network performance.

The training of MLP networks is commonly carried out by the backpropagation algorithm, a ubiquitous method in supervised learning. Backpropagation iteratively adjusts weights and biases to minimize the deviation between predicted and actual outputs. During the training process, the network is exposed to input data and corresponding target outputs. Subsequently, network predictions are compared to targets using a loss function. Gradients of the loss function with respect to network weights and biases are computed and utilized to update these parameters through the gradient descent algorithm. However, MLPs are not without drawbacks. In particular, they show a predisposition to overfitting of training data, and challenges arise when training deep networks with multiple layers because of the vanishing gradient problem.

The RF algorithm (Breiman, 2001) is a sophisticated ensemble forecasting technique that amalgamates multiple individual regression

trees, culminating in a coherent prediction for the target variable. The compositional structure of each regression tree encompasses a root node containing the training dataset, internal nodes establishing conditions on input variables, and leaves denoting the assigned real values for the target variable.

The process of constructing a regression tree involves a stepwise partitioning of the input dataset into subsets, where each subset undergoes the application of a multivariable linear regression model to generate predictions. The iterative progression of the tree is marked by the continual division of each branch into smaller partitions, facilitated by the evaluation of all potential subdivisions within each field. At each iterative step, the subdivision minimizing the least-squared deviation is selected. The least-squared deviation is evaluated using the formula:

$$R(t) = \frac{1}{N(t)} \sum_{i \in t} (y_i - y_m(t))^2 \quad (2)$$

where  $N(t)$  denotes the number of units in the node,  $y_i$  denotes the value of the target variable in the  $i$ -th unit, and  $y_m$  represents the average value of the target variable in node  $t$ .  $R(t)$  is an estimator for the impurity in each node, and the iterative tree-building process persists until the minimum impurity is achieved or another predetermined stopping criterion is met. A pruning procedure is implemented to counteract potential overfitting.

In essence, the RF algorithm encapsulates a dynamic and adaptive approach, harnessing the collective predictive power of individual regression trees. The utilization of multiple trees, each contributing its unique perspective, enriches the predictive capacity of the ensemble model, fostering robustness and resilience against overfitting.

The EN algorithm (Zou & Hastie, 2005) is a prominent tool in machine learning for regression analysis. Distinguished by its seamless integration of two fundamental regression techniques—Lasso and Ridge regression—the EN algorithm emerges as a sophisticated solution, addressing the limitations inherent in its parent methods and providing a robust approach for handling high-dimensional data. Lasso regression introduces a penalty term into the loss function of the linear regression model. This penalty term, expressed as the L1 norm of the coefficient vector, induces the shrinkage of certain coefficients to zero, thereby facilitating feature selection. However, Lasso encounters challenges when confronted with correlated predictors, often selecting only one variable from a group of correlated ones. In contrast, Ridge regression introduces a penalty term into the loss function proportional to the square of the L2 norm of the coefficient vector. While it shrinks all coefficients towards zero, it does not reduce them exactly to zero. Ridge regression handles correlated predictors more effectively than Lasso but does not perform feature selection, thereby posing the risk of overfitting.

The advantage of the EN algorithm lies in its fusion of L1 and L2 penalties into a unified term, harnessing the strengths of both methods. This unique amalgamation empowers EN to adeptly manage correlated predictors and execute feature selection. The regularization strength of EN can be finely tuned by adjusting the hyperparameter alpha, allowing control over the relative influence of the L1 and L2 penalties. EN proves particularly advantageous for datasets with numerous features and a relatively modest number of observations.

## 2.7. Evaluation metrics

The accuracy of the forecast models was assessed using three distinct evaluation metrics: Kling Gupta Efficiency (KGE), RMSE, and Mean Percentage Absolute Error (MAPE). KGE is commonly employed for assessing the accuracy of predictive models in hydrological and hydro-meteorological studies due to its comprehensive evaluation capabilities and robustness. This metric integrates three fundamental statistical measures: Pearson correlation coefficient, variance discrepancy, and mean discrepancy, providing a holistic assessment of the model's performance. RMSE quantifies the average magnitude of the errors between

predicted and observed values by calculating the square root of the average squared differences between predicted and observed values, emphasizing larger errors due to its squared nature. MAPE measures the average percentage difference between predicted and observed values, providing insight into the magnitude of errors relative to the actual values by calculating the absolute percentage difference for each prediction and averaging these values across all predictions, expressing the result as a percentage. Overall, these three metrics provide an effective comprehensive evaluation of the model's accuracy.

In addition, the distribution of the Relative Error (RE) has been analyzed. Table 4 provides a comprehensive description of the evaluation metrics along with their corresponding equations.

## 2.8. Model development

To achieve a multi-step ahead prediction of ETo, the predictors include some lagged values of ETo, in addition to the current day's ETo and exogenous climatic variables. Notably, for the sake of simplifying

**Table 4**  
Evaluation metrics.

<b>Kling Gupta Efficiency</b>	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_p}{\sigma_a} - 1\right)^2 + \left(\frac{\overline{ETo_p}}{\overline{ETo_a}} - 1\right)^2}$	(3)
Assesses the overall accuracy of a model's predictions, taking into account the correlation, bias, and variability between actual and predicted values in hydrological simulations. It provides a comprehensive measure of the model's ability to capture key characteristics of the observed hydrograph, including timing, shape, and magnitude. Range: $-\infty - 1$ , with higher values indicating more accurate models.		
<b>Root Mean Square Error</b>	$RMSE = \sqrt{\frac{\sum_{i=1}^n (ETo_p^i - ETo_a^i)^2}{n}}$	(4)
The square root of the total squared error between the predicted and measured ETo, divided by the number of samples and then normalized. Range: $0 - \infty$ , with lower values indicating more accurate models.		
<b>Mean Absolute Percentage Error</b>	$MAPE = \frac{\sum_{i=1}^n \left  \frac{ETo_p^i - ETo_a^i}{ETo_a^i} \right }{n}$	(5)
The relative error between the predicted and actual ETo, normalized by the number of samples. Range: $0 - \infty$ , with lower values indicating more accurate models.		
<b>Relative error</b>	$Relative\ error = \frac{ETo_p^i - ETo_a^i}{ETo_a^i}$	(6)
The ratio between the difference of predicted and actual ETo and the actual ETo. Range: $0 - \infty$ , with lower values indicating more accurate models.		
where $ETo_a^i$ = actual evapotranspiration for the $i^{th}$ data and $ETo_p^i$ = predicted reference evapotranspiration for the $i^{th}$ data, $\overline{ETo_a}$ = mean of the actual reference evapotranspiration, $\overline{ETo_p}$ = mean of the predicted reference evapotranspiration, $r$ = correlation coefficient between actual and predicted reference evapotranspiration, $\sigma_p$ = standard deviation of the predicted reference evapotranspiration, $\sigma_a$ = standard deviation of the actual reference evapotranspiration, $n$ = number of data vectors.		

and expediting model development, lagged values of climatic variables were excluded from the input variables. It is crucial to emphasize that, despite this exclusion, the lagged values of ETo implicitly encompass these climatic variables. The optimal number of lagged values, denoted as  $n$ , and the hyperparameters of the models were determined through a Bayesian Optimization (BO) procedure. Bayesian optimization is a powerful method for optimizing functions that pose challenges in evaluation and incur significant costs. This approach adopts a sequential model-based strategy, leveraging a probabilistic model to grasp the characteristics of the function under optimization. As fresh data points emerge, the model undergoes progressive refinement via Bayesian inference, assimilating prior knowledge about the function and uncertainties surrounding the model. This algorithm adeptly navigates between exploring new territories within the search space and exploiting areas likely to yield optimal performance. Consequently, it uncovers the global optimum with fewer objective function evaluations compared to alternative optimization techniques. While a detailed procedure description is omitted here for brevity, interested readers can delve into essential literature for deeper insights (e.g., Snoek et al., 2012). In this study, the Root Mean Square Error (RMSE) was selected as the metric to be minimized. The multi-step ahead forecast was executed using a hybrid direct-recursive procedure. In this approach, a distinct model is formulated for each time step to be predicted. However, each model utilizes predictions from prior time steps as new lagged ETo values to forecast future ETo values. Given the unknown nature of future values during the forecasting process, the previously predicted values serve as a reasonable estimate for the actual values. The models were trained using 70 % of the time series data spanning from January 1, 1981, to May 26, 2010, while the remaining 30 % was reserved for testing, covering the period from May 27, 2010, to December 31, 2022. A preliminary analysis revealed that this partitioning led to optimal results, while further increasing the size of the training set did not yield any significant improvements in model accuracy. Recognizing the temporal essence of the prediction challenge, it was imperative to maintain temporal coherence within the training intervals. To ensure comparability and consistency, the data inputs were normalized using the following equation:

$$x_{t\_norm} = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (7)$$

where  $x_t$ ,  $x_{min}$ , and  $x_{max}$  are, respectively, the values of the variable  $x$  at time  $t$ , the minimum value, and the maximum value.

## 3. Results

Table 5 displays the evaluation metrics, with reference to the testing phase, for various forecasting horizons, different combinations of input variables, and various considered algorithms. As an illustrative example, Figs. 3 and 4 delineate the comparison between forecasted and actual ETo values for all analyzed combinations of input variables and across all considered algorithms, corresponding to forecasting horizons of 7 days and 60 days, respectively. Fig. 5, conversely, depicts the time series of predicted and actual ETo values pertaining to models that incorporate all exogenous input variables with a forecasting horizon of 30 days.

With reference to the 7-day forecasting horizon, models incorporating all input variables significantly outperformed both attribute-based models obtained from PCA and those based on features selected through PSO. Regarding algorithms, the MLP-RF Stacked Model demonstrated markedly higher accuracy than RBF, which, in turn, outperformed XNV when considering all variables. For models based on PCA attribute or PSO-selected features, XNV exhibited lower values in both RMSE and MAPE compared to RBF. Furthermore, the MLP-RF Stacked Model based on all input variables achieved excellent evaluation metric values, specifically  $KGE = 0.998$ ,  $RMSE = 0.0904$  mm/day, and  $MAPE = 3.97\%$ .

Extending the forecasting horizon to 14 days has revealed a modest

**Table 5**

Summary of results in terms of evaluation metrics with reference to the testing phase.

Forecasting horizon	Predictors	RBF			XNV			MLP-RF Stacked Model		
		KGE	RMSE [mm/day]	MAPE [%]	KGE	RMSE [mm/day]	MAPE [%]	KGE	RMSE [mm/day]	MAPE [%]
7 days ahead	All input variables	0.985	0.1565	6.335	0.940	0.1689	7.337	0.998	0.0904	3.965
	PCA attributes	0.969	0.2523	10.356	0.924	0.2082	9.177	0.989	0.1225	5.416
	Features selection using PSO	0.960	0.2435	9.287	0.968	0.2265	8.651	0.985	0.1628	6.904
14 days ahead	All input variables	0.983	0.1615	6.663	0.936	0.1740	7.871	0.997	0.1019	4.123
	PCA attributes	0.964	0.2645	10.952	0.917	0.2171	10.113	0.988	0.1244	5.595
	Features selection using PSO	0.953	0.2546	9.879	0.966	0.2342	9.039	0.982	0.2081	8.060
30 days ahead	All input variables	0.983	0.1627	6.785	0.935	0.1753	8.066	0.997	0.1023	4.159
	PCA attributes	0.962	0.2695	11.289	0.914	0.2198	10.442	0.987	0.1248	5.655
	Features selection using PSO	0.950	0.2603	10.228	0.965	0.2383	9.277	0.980	0.2115	8.291
60 days ahead	All input variables	0.983	0.1627	6.814	0.935	0.1755	8.101	0.997	0.1025	4.177
	PCA attributes	0.962	0.2705	11.382	0.913	0.2203	10.600	0.987	0.1250	5.686
	Features selection using PSO	0.950	0.2616	10.322	0.965	0.2395	9.357	0.980	0.2121	8.356

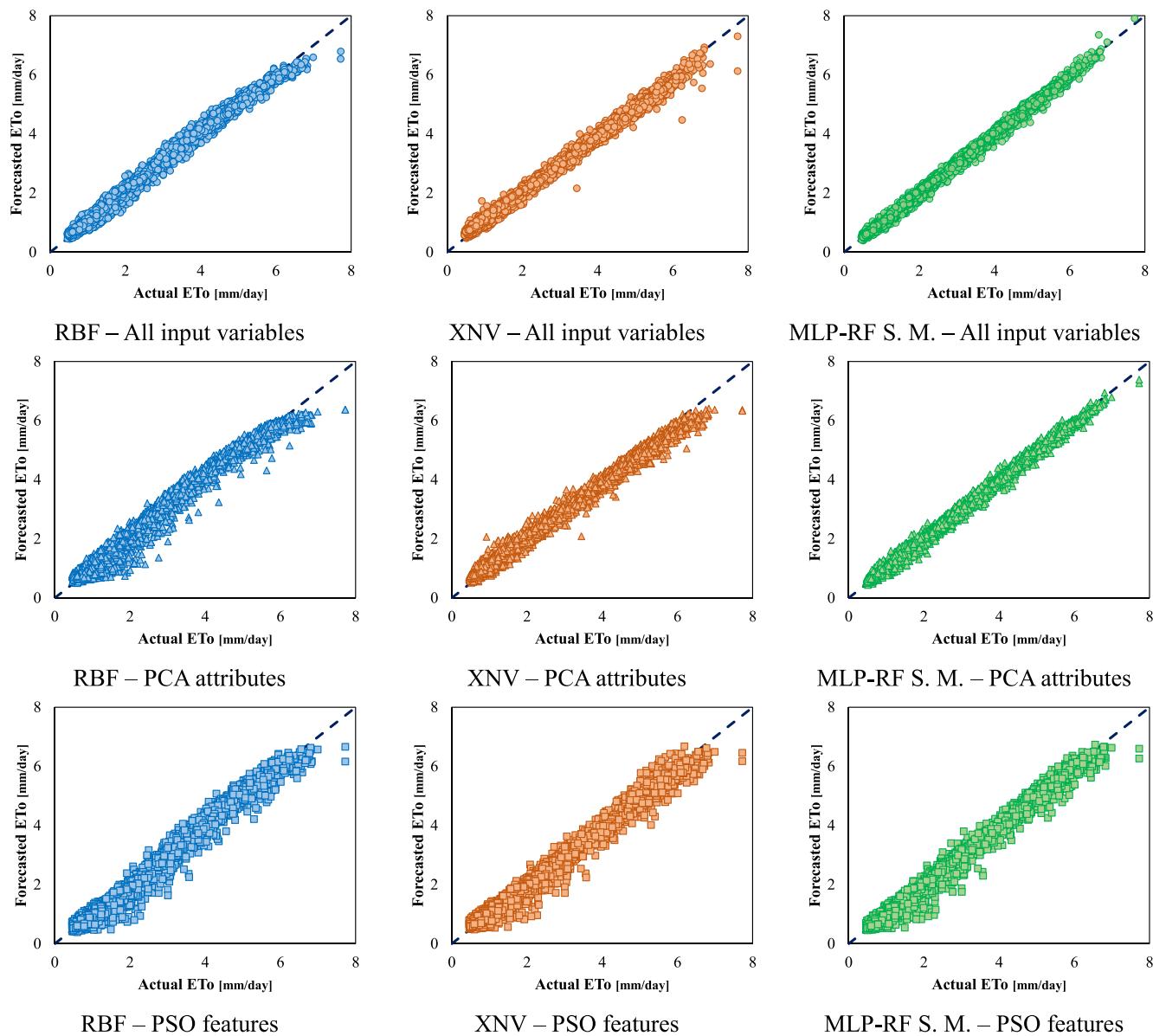


Fig. 3. Forecast values versus actual ETo values for a 7-day forecast horizon.

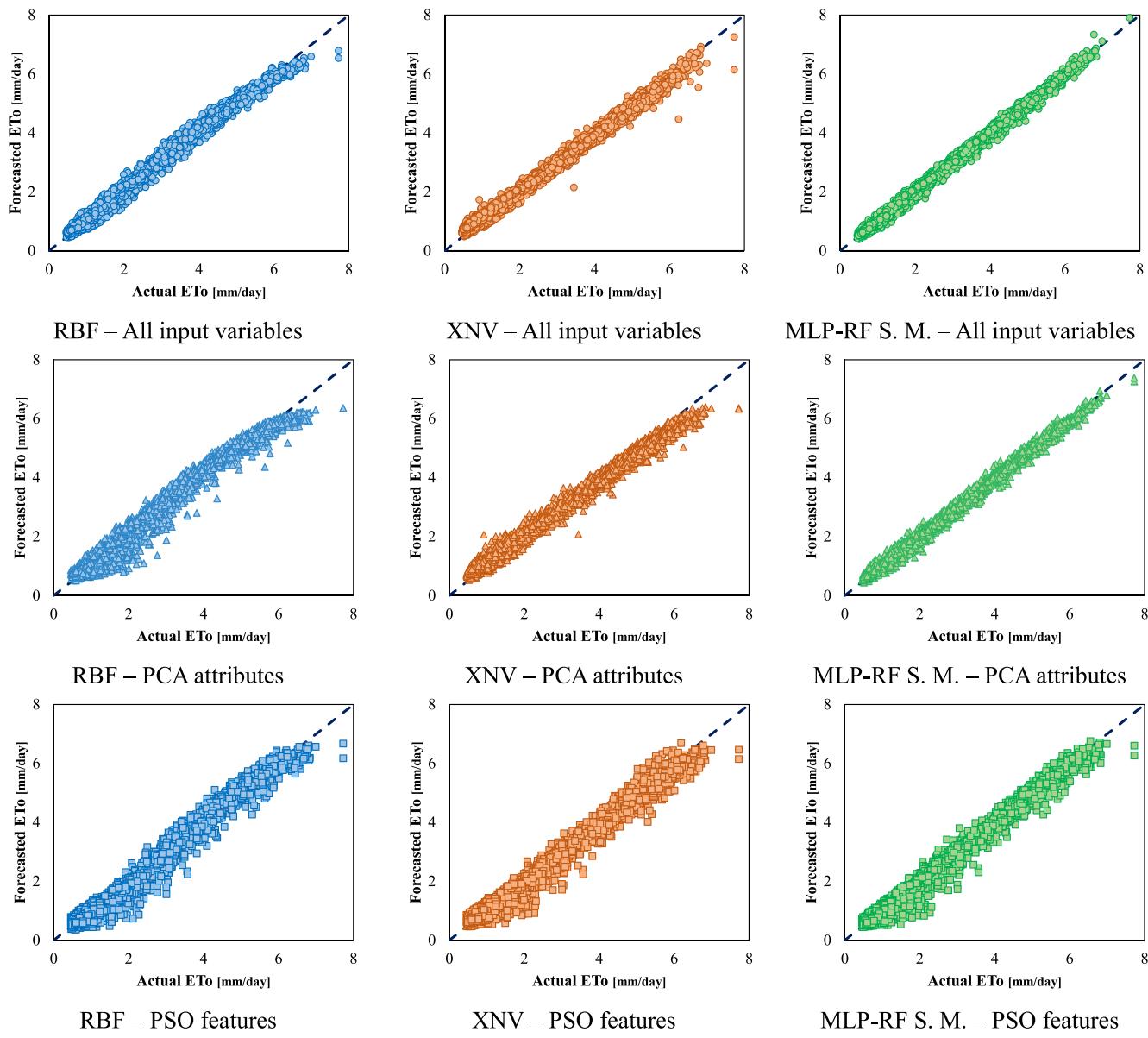


Fig. 4. Forecast values versus actual ET<sub>0</sub> values for a 60-day forecast horizon.

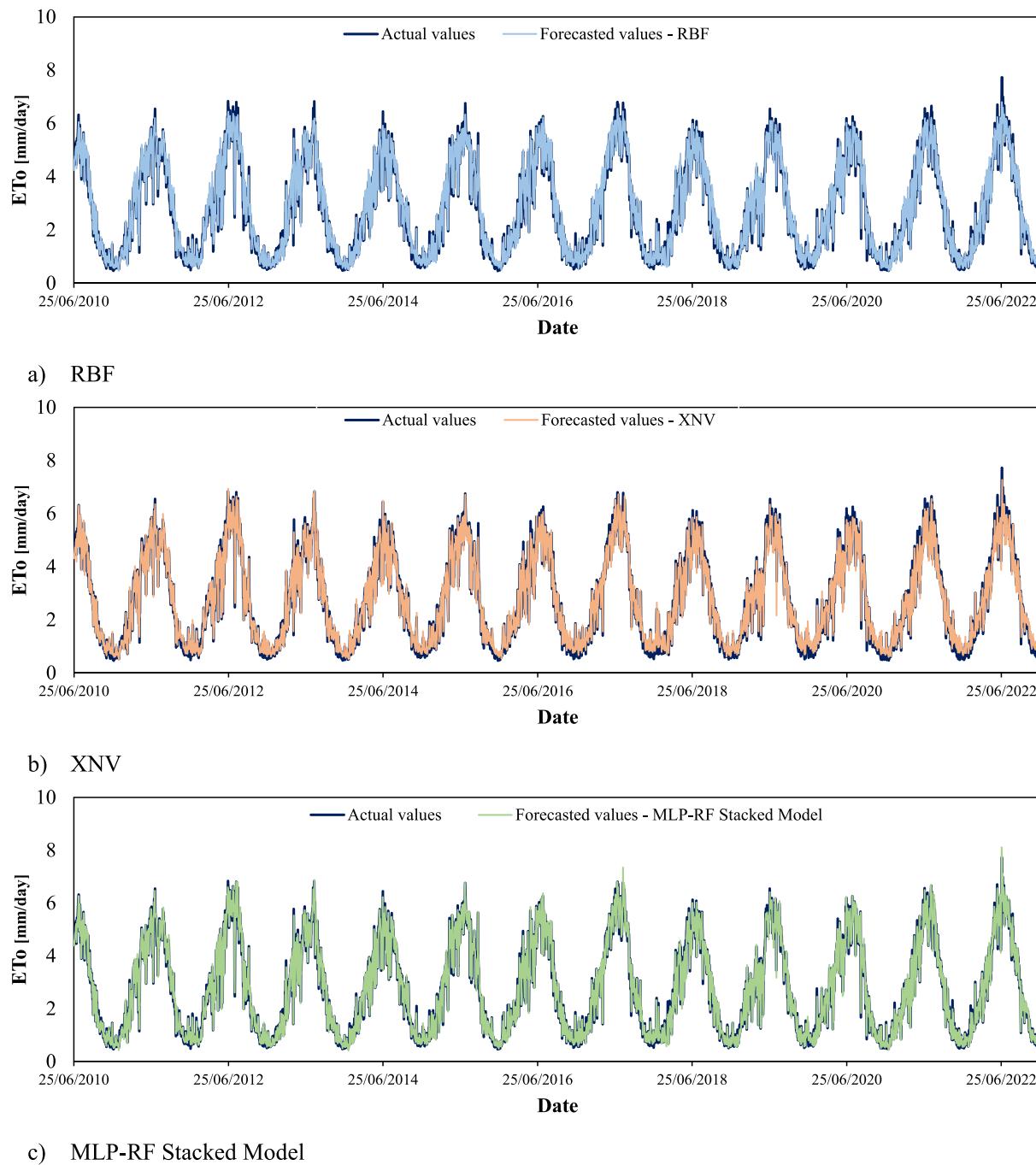
decrease in predictive accuracy.

The MLP-RF Stacked Model has once again demonstrated itself to be the most accurate algorithm for all combinations of input variables, clearly outperforming XNV. In turn, XNV exhibited lower forecasting errors compared to RBF but also generally lower values of KGE.

Regarding models based on all input variables, KGE has undergone nearly negligible deteriorations, within the range of 0.1 % to 0.43 %, while the increase in RMSE has ranged from 3 % to 12.7 %, and that of MAPE from 3.98 % to 7.28 %. The MLP-RF Stacked Model experienced the highest percentage increase in RMSE, whereas XNV exhibited the highest percentage increase in MAPE. As for models based on attributes defined through PCA, KGE has experienced deteriorations within the range of 0.1 % to 0.76 %, RMSE has incurred deteriorations ranging from 1.55 % to 4.84 %, while the increase in MAPE has fallen within the range of 3.31 % to 10.20 %. The highest percentage increase in RMSE was computed for RBF, whereas XNV exhibited the highest percentage increase in MAPE. Regarding models based on features selected through PSO, KGE has undergone reductions within the range of 0.21 % to 0.73 %, while increases in RMSE have been calculated within the range of 3.40 % to 27.83 %. For MAPE, increments have been observed within

the range of 4.34 % to 16.74 %. The Stacked Model exhibited the highest percentage increases in both RMSE and MAPE. However, it has consistently demonstrated itself as the most accurate forecasting algorithm among the three considered, even for the 14-day forecasting horizon. On the other hand, both RBF and XNV have demonstrated less sensitivity to the reduction in the number of input variables.

Extending the forecasting horizon to 30 days has resulted in only minimal reductions in accuracy compared to the 14-day forecasts. Models incorporating all input variables still exhibit excellent metric values, particularly the MLP-RF Stacked Model (KGE = 0.997, RMSE = 0.1023 mm/day, MAPE = 4.16 %), outperforming RBF, which in turn outperformed XNV. As observed in the predictive horizons of 7 and 14 days, in the case of the 30-day forecasting horizon, the model incorporating all input variables distinctly outperforms, under identical algorithms, those based on PCA attributes and features selected through PSO. Nevertheless, the latter approaches still yield highly accurate predictions. Furthermore, transitioning from 14-day to 30-day forecasts, models incorporating all variables experienced a negligible reduction in KGE, an increase in RMSE within the range of 0.39–0.75 %, and an increase in MAPE within the range of 0.87–2.48 %. For models based on



**Fig. 5.** Time series of actual and predicted ETo, with a 30-day forecast horizon. All input variables.

PCA-defined attributes, a reduction in KGE within the range of 0.1–0.33 %, an increase in RMSE within the range of 0.32–1.89 %, and an increase in MAPE within the range of 1.07–3.25 % were observed. Lastly, for models based on the limited set of variables identified through PSO, the reduction in KGE fell within the range of 0.1–0.31 %, the increase in RMSE ranged from 1.63 % to 2.24 %, while the increase in MAPE ranged from 2.63 % to 3.53 %. The MLP-RF Stacked Model exhibited the least reduction in accuracy compared to models incorporating all input variables and those based on attributes defined by PCA. Similarly, XNV demonstrated the least decrease in accuracy among models based on the limited set of features selected through PSO.

Finally, concerning the forecast horizon extending to 60 days ahead, the accuracy of the models was only marginally lower than that associated with the 30-day forecasts. All considered models provided

exceptionally accurate predictions, with the most precise values achieved through the MLP-RF Stacked Model trained with all input variables (KGE = 0.997, RMSE = 0.103 mm/day, MAPE = 4.18 %).

A direct and immediate comparison of the accuracies of different algorithms for each input variable scenario and for each forecasting horizon can also be achieved through Taylor diagrams, as proposed in Fig. 6. Taylor diagrams were formulated by leveraging the geometric interplay among the correlation coefficient ( $r$ ), standard deviation, and RMSE (usually indicated as Root Mean Square Differences, RMSD, in the Taylor diagrams analysis). The Taylor diagrams accentuate the precision and effectiveness of the examined models in comparison to the observed values. Owing to its efficacy, the Taylor diagram has garnered extensive adoption in hydrological investigations (Pham et al., 2022).

The analysis of Taylor diagrams confirmed what was already

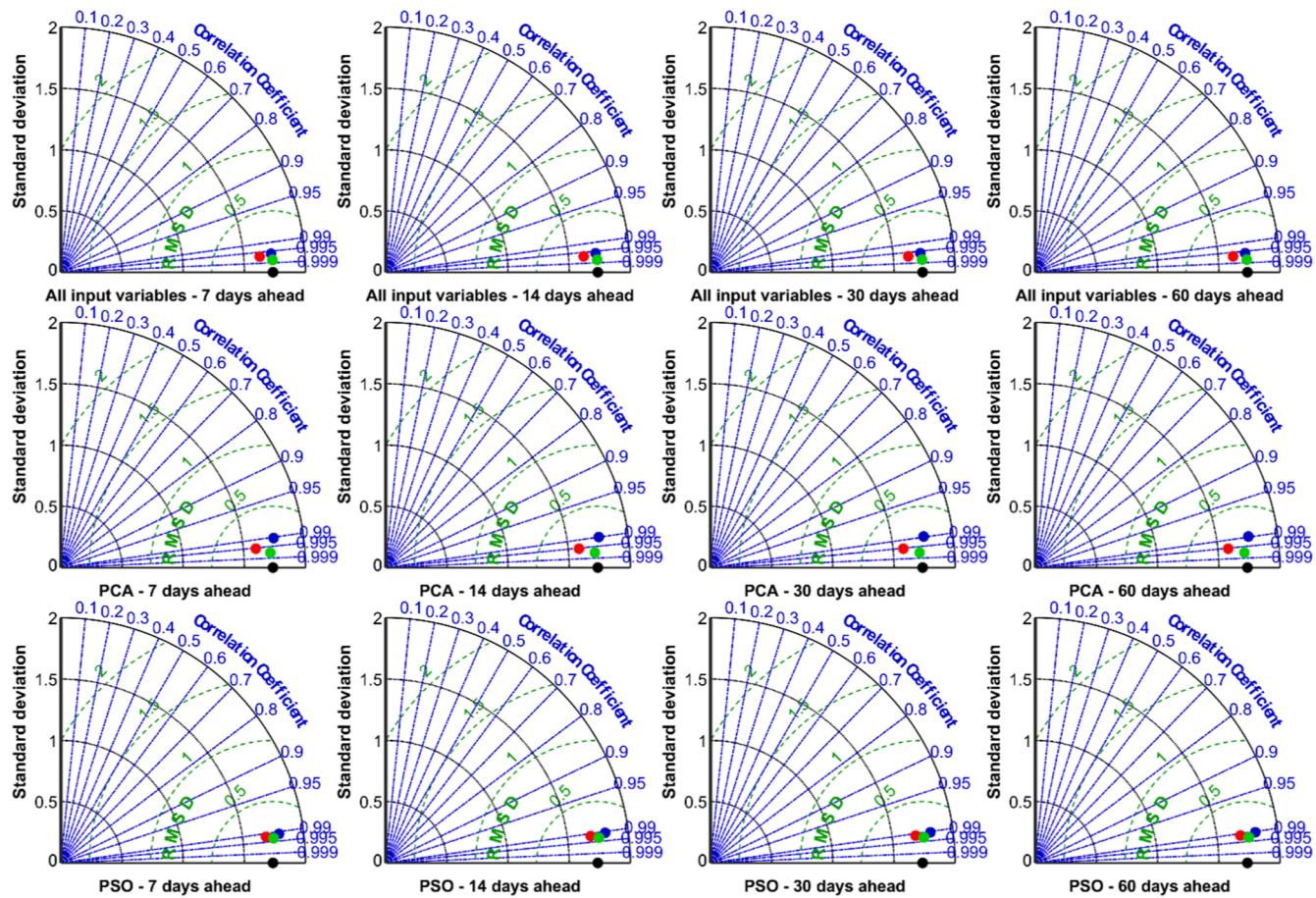


Fig. 6. Taylor diagrams for all models and forecast horizons.

observed from the evaluation metrics analysis, with the MLP-RF Stacked Model emerging as the most accurate algorithm in all investigated circumstances. Conversely, RBF and XNV yielded comparable results in most cases, with XNV being preferable in scenarios with a reduced number of input variables selected through PSO.

Additional insight into the accuracy of the models here developed was gained through an examination of the combined violin-box plots depicted in Fig. 7. These plots illustrate the Relative Error of various models across different forecasting horizons.

In particular, from Fig. 7a, pertaining to models that incorporate all input variables, it is evident that the MLP-RF Stacked Model exhibited a negligible bias and the smallest Interquartile Range (IQR) values, characterized by a negligible increase with the forecasting horizon. Slightly higher IQRs were observed in the case of RBF, which also displayed a slight bias and a smaller number of outliers. XNV, on the other hand, exhibited a noticeably more pronounced bias, along with the highest IQR values, increasing with the forecasting horizon. Across all algorithms, the majority of outliers were positive (overestimation), primarily affecting lower ET<sub>0</sub> values.

From Fig. 7b, pertaining to models based on attributes obtained through PCA, it is evident that the MLP-RF Stacked Model, as observed in the analysis of other evaluation metrics, was markedly the most accurate. It exhibited nearly negligible biases and IQR values that were quite small and comparable to those obtained in the case of models incorporating all input variables. In this context, RBF and XNV demonstrated fairly similar performances, with RBF characterized by slightly lower bias and IQR values but also featuring a more significant number of outliers, both positive and negative.

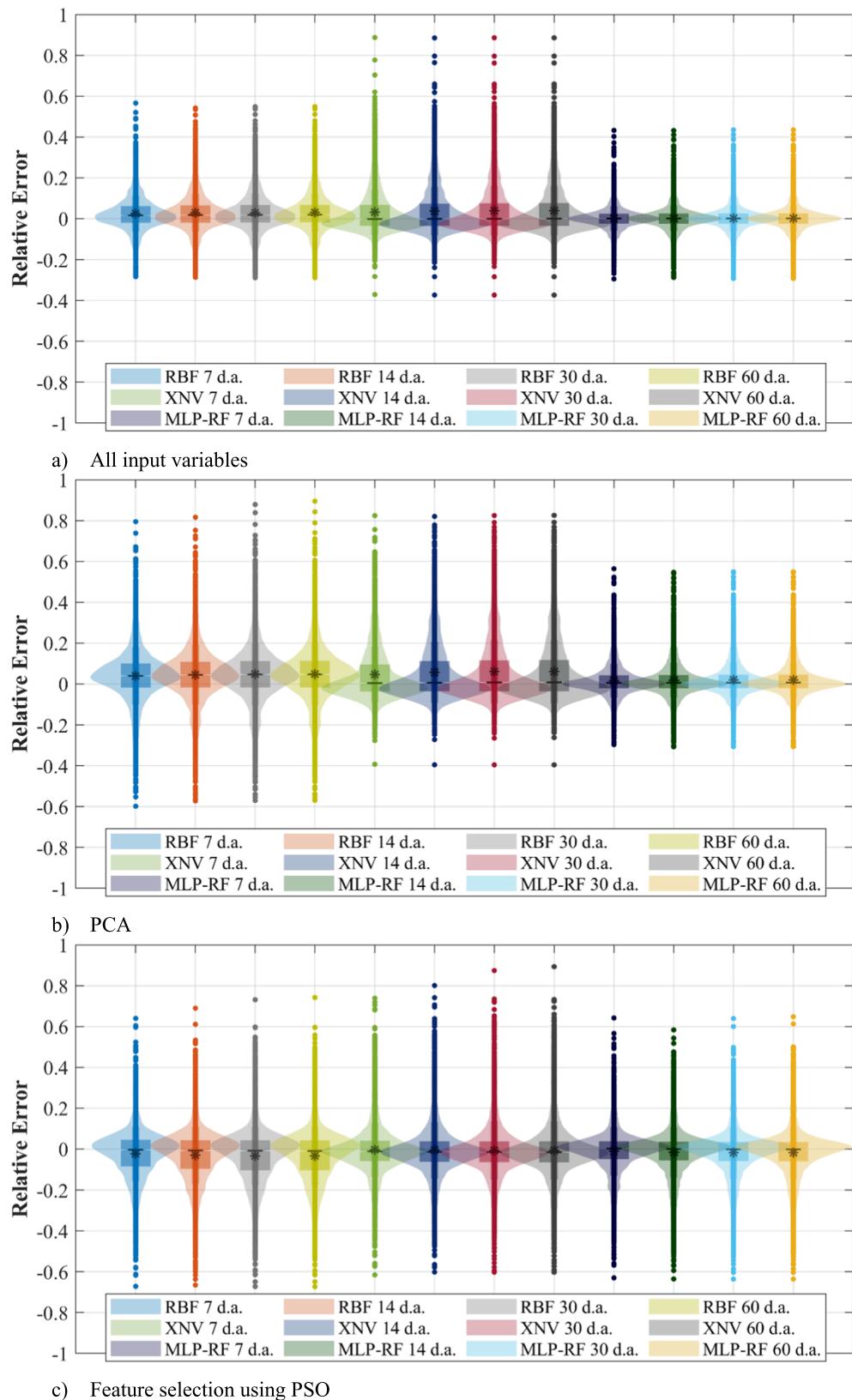
Fig. 7c, pertaining to models that need as inputs only T<sub>mean</sub>, RH<sub>max</sub>, and Rs, reveals that in this case, XNV demonstrated predictive accuracy

slightly below that of the MLP-RF Stacked Model, with comparable bias values and slightly higher IQR values. XNV significantly outperformed RBF, which, in turn, exhibited higher bias and IQR values.

#### 4. Discussion

The outcomes analyzed in the previous section reveal that the three algorithms under consideration facilitate the development of highly accurate predictive models for medium-term forecasting horizons, even when accounting solely for limited climatic variables such as T<sub>mean</sub>, RH<sub>max</sub>, and Rs. This achievement is made possible by the availability of an exceptionally extensive training dataset, encompassing a very long time series of all relevant variables. Furthermore, the predictability is enhanced by the cyclic nature of the climatic parameter ET<sub>0</sub>. The elevated predictive performance is chiefly ensured by the MLP-RF Stacked Model but is also observed in the case of RBF and XNV.

The overwhelming majority of studies assessing evapotranspiration through machine learning models based on a limited number of climatic variables focus on estimating the current value (e.g., Laaboudi et al., 2012; Wen et al., 2015; Granata, 2019; Chen et al., 2020; Ferreira and Da Cunha, 2020b; Sharma et al., 2022). However, the literature also presents examples of studies aimed at predicting future values, which can serve as benchmarks for the results obtained here. Torres et al. (2011) employed models based on a Multivariate Relevance Vector Machine (MVRVM) or MLP to forecast ET<sub>0</sub> in Central Utah with a forecasting horizon ranging from 1 to 7 days. For the 7-day predictions, they achieved a minimum RMSE of 0.85 mm/day, a value significantly higher than the one obtained in this study. It is noteworthy that the authors considered 7 or 10 past values of ET<sub>0</sub> and air temperature as the sole exogenous variables in their input parameters. Traore et al. (2016)



**Fig. 7.** Combined violin-box plots of relative errors in all models considered and for all forecast horizons.

employed four distinct Artificial Neural Network (ANN) models for forecasting up to 15 days of ET<sub>0</sub> in irrigated areas near Dallas, Texas, USA. The authors considered three sets of input combinations drawn from T<sub>min</sub>, T<sub>max</sub>, Rs, and extraterrestrial solar radiation (R<sub>a</sub>), achieving

optimal results with the MLP algorithm. This algorithm resulted in a mean square error of 0.770 mm/day, significantly higher than the errors estimated for the algorithms utilized in this study. Ferreira & Da Cunha (2020a) employed models based on LSTM, 1D CNN, and a combination

of the two algorithms, along with more traditional machine learning algorithms such as ANN and RF, for forecasting daily ETo up to 7 days in various regions of Minas Gerais, Brazil. Considering different combinations of input variables, the most comprehensive of which included a number of lagged values of ETo and climatic variables such as  $T_{\max}$ ,  $T_{\min}$ , RH<sub>max</sub>, RH<sub>min</sub>, WS<sub>10</sub>, net solar radiation (RN), and R<sub>a</sub>, the authors obtained an average RMSE value of 0.88 mm/day, significantly higher than that achieved in the present study. [Mandal & Chanda \(2023\)](#) employed six distinct machine learning algorithms, namely Support Vector Regression (SVR), Multivariate Adaptive Regression Splines (MARS), RF, MLP, 1D CNN, and LSTM, for forecasting ETo up to 28 days ahead across various agroclimatic zones of India. Similar to the present study, these authors utilized ERA5 data and obtained results exhibiting accuracy comparable to or slightly lower than those reported in the current investigation. Additionally, these authors observed a modest decline in accuracy with the extension of the forecasting horizon, consistent with the findings presented herein.

The MLP-RF Stacked Model employed in this study, recently successful in forecasting streamflow ([Granata et al., 2022](#)), soil water content ([Granata et al. 2023](#)), lake water temperature ([Di Nunno et al., 2023a](#)), and groundwater levels ([Di Nunno et al., 2023b](#)), reaffirms itself as a highly effective predictive tool for hydrological variables, including ETo. On the other hand, a Stacked Model is a specific type of ensemble model, a category of ML models that have already proven effective in the multi-step ahead prediction of ETo ([Nourani et al., 2020](#)). Stacked machine learning models combine predictions from diverse base learners via a *meta*-learner, capitalizing on their individual strengths to enhance predictive accuracy. Their versatility allows for the integration of domain knowledge and feature engineering, facilitating the capture of intricate data patterns. However, their implementation entails challenges, including heightened complexity, computational demands, and vulnerability to overfitting if not judiciously regularized. On the other hand, the XNV algorithm, unused so far in hydrological applications, is distinguished by its ability to efficiently handle high-dimensional data and nonlinear relationships, making it particularly well-suited for complex pattern recognition tasks. Additionally, the algorithm is known for its robustness to noisy data and its capability to provide interpretable models, aiding in the understanding of underlying data patterns. These features make it an effective alternative to many better-known and commonly used algorithms. There is no shortage of negative features, primarily the high sensitivity to the choice of hyperparameters.

Feature selection procedures confirm their crucial role in improving the interpretability of predictive models in machine learning while retaining high efficiency and accuracy. Their foremost advantage lies in the capability of reducing the dimensionality of datasets while retaining essential information. By transforming correlated variables into linearly uncorrelated components, PCA simplifies the computational burden and enhances model interpretability, facilitating better insights into underlying patterns. Moreover, PCA aids in identifying latent features that contribute most significantly to variance, thereby mitigating the risk of overfitting. However, PCA also presents drawbacks. One notable limitation is the assumption of linearity, which may not hold true for complex, nonlinear relationships inherent in many real-world datasets. Consequently, PCA might inadequately capture the intrinsic structure of the data, leading to suboptimal performance in predictive tasks. Furthermore, PCA operates under the assumption of Gaussian distributions, which might not be applicable in all scenarios, potentially compromising its efficacy in capturing the true underlying data distribution.

PSO efficiently explores the feature space, adapting dynamically to find optimal subsets and avoiding local optima. PSO's flexibility in defining objective functions tailored to ETo modeling objectives enhances its utility. However, challenges may arise in noisy or redundant feature scenarios, requiring careful parameter tuning. It is noteworthy to observe that the use of PSO has underscored the significance of the exogenous variable relative humidity in multi-step ahead forecasts of

ETo. This finding is consistent with recent results obtained by [Elzain et al. \(2024\)](#).

In many cases, a reduction in variables is also expected to lead to an improvement in model accuracy. This could occur when the overall set of input variables does not adequately represent the studied phenomenon, or in cases of overfitting. However, these scenarios did not impact the study conducted here.

The forecasting models presented can bring about numerous tangible benefits, particularly for irrigation service managers and the farming community. Accurate ETo prediction models provide a basis for precise irrigation scheduling, enabling farmers to align water application with the actual water demand of crops. This optimization minimizes water wastage, reduces energy consumption, and contributes to the conservation of water resources. Precision irrigation, guided by ETo models, ensures that crops receive the optimal amount of water needed for their growth stages. This targeted approach promotes higher crop yields, improved quality, and ultimately contributes to increased agricultural productivity. Accurate ETo models facilitate the judicious use of water resources, contributing to overall resource conservation. By avoiding over-irrigation, farmers can reduce runoff, leaching, and nutrient loss, thereby promoting environmental sustainability. Precision agriculture allows for the efficient use of inputs such as water, fertilizers, and energy. This targeted resource allocation minimizes unnecessary costs, making agriculture more economically viable for farmers. In addition, in the face of climate variability, precise ETo models provide a valuable tool for adapting irrigation practices to changing conditions. This adaptability is essential for mitigating the impact of climate change on agricultural productivity.

In light of the foregoing discussion, the ETo prediction emerges as a cornerstone in scientific pursuits related to water management, agriculture, and environmental sustainability. Accurate forecasts of ETo empower stakeholders to implement proactive measures that enhance resource efficiency, reduce environmental impact, and contribute to the resilience of ecosystems in the face of evolving climatic patterns. As we navigate the challenges of a changing world, the meticulous consideration of ETo remains an invaluable tool in our quest for sustainable and responsible water stewardship.

The primary limitation of this study lies in its focus on a singular study area characterized by a typical Mediterranean climate. The predictive accuracy may be adversely influenced by conditions marked by broader and irregular fluctuations in climatic variables. This aspect will be investigated in future research endeavours, which will also be directed towards exploring the effectiveness of the introduced models for long-term forecasting horizons.

## 5. Conclusions

This study is grounded in the application of two innovative algorithms, the MLP-RF Stacked Model and the XNV, to forecast ETo up to a medium-term forecast horizon of 60 days in Agro Pontino. Situated in Central Italy, Agro Pontino is renowned for its vibrant agricultural production within the broader Mediterranean Europe. A third algorithm, the RBF-NN, was employed as a benchmark. Three distinct sets of input variables were considered. The key findings are as follows:

- The models based on the complete set of exogenous climatic variables have proven to be the most accurate. Nevertheless, even those requiring only mean temperature, maximum relative humidity, and shortwave solar radiation as input have yielded very good results, even with regard to 60-day forecasting horizons.
- The models based on MLP-RF Stacked Model have outperformed those based on XNV and RBF for all combinations of input variables and forecasting horizons. XNV outperformed RBF when considering only mean temperature, maximum humidity, and shortwave solar radiation as exogenous input variables.

- The reduction in accuracy with the extension of the forecasting horizon is notably mild. This circumstance suggests the possibility of achieving accurate results even for significantly longer forecasting horizons with the developed models.

The presented forecasting models offer significant advantages for irrigation management and agriculture. Accurate ETo predictions enable precise irrigation scheduling, reducing water wastage, and conserving resources. This targeted approach enhances crop yields, quality, and environmental sustainability, making agriculture more economically viable and adaptable to climate variability.

## CRediT authorship contribution statement

**Francesco Granata:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Fabio Di Nunno:** Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Data curation. **Giovanni de Marinis:** Writing – review & editing, Validation, Investigation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The utilized data are freely available online.

## References

- Aghelpour, P., Norooz-Valashedi, R., 2022. Predicting daily reference evapotranspiration rates in a humid region, comparison of seven various data-based predictor models. *Stoch. Env. Res. Risk A* 36, 4133–4155.
- 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, 300.
- Bai, Y., Zhang, S., Bhattacharai, N., Mallik, K., Liu, Q., Tang, L., Zhang, J., 2021. On the use of machine learning based ensemble approaches to improve evapotranspiration estimates from croplands across a wide environmental gradient. *Agric. For. Meteorol.* 298, 108308.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Chen, Z., Zhu, Z., Jiang, H., Sun, S., 2020a. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *J. Hydrol.* 591, 125286.
- Chen, Z., Sun, S., Wang, Y., Wang, Q., Zhang, X., 2020b. Temporal convolution-network-based models for modeling maize evapotranspiration under mulched drip irrigation. *Comput. Electron. Agric.* 169, 105206.
- Chia, M.Y., Huang, Y.F., Koo, C.H., 2020. Support vector machine enhanced empirical reference evapotranspiration estimation with limited meteorological parameters. *Comput. Electron. Agric.* 175, 105577.
- Cioffi, F., Conticello, F., Lall, U., Marotta, L., Telesca, V., 2017. Large scale climate and rainfall seasonality in a Mediterranean area: insights from a non-homogeneous Markov model applied to the agro-pontino plain. *Hydrol. Process.* 31, 668–686.
- Di Nunno, F., Zhu, S., Ptak, M., Sojka, M., Granata, F., 2023a. A stacked machine learning model for multi-step ahead prediction of lake surface water temperature. *Sci. Total Environ.* 890, 164323.
- Di Nunno, F., Giudicianni, C., Creaco, E., Granata, F., 2023b. Multi-step ahead groundwater level forecasting in grand Est, France: Comparison between stacked machine learning model and radial basis function neural network. *Groundw. Sustain. Dev.* 23, 101042.
- Dou, X., Yang, Y., 2018. Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Comput. Electron. Agric.* 148, 95–106.
- Elzain, H.E., Abdalla, O.A., Abdallah, M., Al-Maktoumi, A., Eltayeb, M., Abba, S.I., 2024. Innovative approach for predicting daily reference evapotranspiration using improved shallow and deep learning models in a coastal region: a comparative study. *J. Environ. Manage.* 354, 120246.
- Feng, Y., Peng, Y., Cui, N., Gong, D., Zhang, K., 2017. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.* 136, 71–78.
- Ferreira, L.B., da Cunha, F.F., 2020. 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.
- Gocić, M., Motamed, S., Shamshirband, S., Petković, D., Ch, S., Hashim, R., Arif, M., 2015. Soft computing approaches for forecasting reference evapotranspiration. *Comput. Electron. Agric.* 113, 164–173.
- 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.
- Granata, F., 2019. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agric Water Manag* 217, 303–315.
- Granata, F., Di Nunno, F., 2021. Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. *Agric Water Manag* 255, 107040.
- Granata, F., Di Nunno, F., de Marinis, G., 2022. Stacked machine learning algorithms and bidirectional long short-term memory networks for multi-step ahead streamflow forecasting: a comparative study. *J. Hydrol.* 613, 128431.
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24 (6), 417.
- Jia, W., Zhang, Y., Wei, Z., Zheng, Z., Xie, P., 2023. Daily reference evapotranspiration prediction for irrigation scheduling decisions based on the hybrid PSO-LSTM model. *PLoS One* 18 (4), e0281478.
- Kennedy, J., Eberhart, R., 1995. Particle Swarm Optimization. In: Proceedings of ICNN'95-International Conference on Neural Networks. IEEE, pp. 1942–1948.
- Kisi, O., Alizamir, M., 2018. Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: wavelet extreme learning machine vs wavelet neural networks. *Agric. For. Meteorol.* 263, 41–48.
- Laaboudi, A., Mouhouche, B., Draoui, B., 2012. Neural network approach to reference evapotranspiration modeling from limited climatic data in arid regions. *Int. J. Biometeorol.* 56 (5), 831–841.
- Landeras, G., Ortiz-Barredo, A., López, J.J., 2009. Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. *J. Irrig. Drain. Eng.* 135 (3), 323–334.
- Luo, Y., Traore, S., Lyu, X., Wang, W., Wang, Y., Xie, Y., Jiao, X., Fipps, G., 2015. Medium range daily reference evapotranspiration forecasting by using ANN and public weather forecasts. *Water Resour. Manag.* 29 (10), 3863–3876.
- Malik, A., Kumar, A., Ghorbani, M.A., Kashani, M.H., Kisi, O., Kim, S., 2019. The viability of co-active fuzzy inference system model for monthly reference evapotranspiration estimation: case study of Uttarakhand state. *Hydrol. Res.* 50, 1623–1644. <https://doi.org/10.2166/nh.2019.059>.
- Mandal, N., Chanda, K., 2023. Performance of machine learning algorithms for multi-step ahead prediction of reference evapotranspiration across various agro-climatic zones and cropping seasons. *J. Hydrol.* 620, 129418.
- McWilliams, B., Balduzzi, D., Buhmann, J.M., 2013. Correlated random features for fast semi-supervised learning. *Adv. Neural Inf. Proces. Syst.* 26.
- Murtagh, F., 1991. Multilayer perceptrons for classification and regression. *Neurocomputing* 2 (5–6), 183–197.
- Nourani, V., Elkiran, G., Abdullahi, J., 2020. Multi-step ahead modeling of reference evapotranspiration using a multi-model approach. *J. Hydrol.* 581, 124434.
- Parisse, B., Alilla, R., Pepe, A.G., De Natale, F., 2023. MADIA - meteorological variables for agriculture: a dataset for the Italian area. *Data Brief* 46, 108843.
- Park, J., Sandberg, I.W., 1991. Universal approximation using radial-basis-function networks. *Neural Comput.* 3 (2), 246–257.
- Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11), 559–572.
- Pham, Q.B., Kumar, M., et al., 2022. Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Comput. & Applic.* 34, 10751–10773. <https://doi.org/10.1007/s00521-022-07009-7>.
- Sharma, G., Singh, A., Jain, S., 2022. A hybrid deep neural network approach to estimate reference evapotranspiration using limited climate data. *Neural Comput. & Applic.* 1–20.
- Tikhamarine, Y., Malik, A., Kumar, A., Souag-Gamane, D., Kisi, O., 2019. Estimation of monthly reference evapotranspiration using novel hybrid machine learning approaches. *Hydrol. Sci. J.* 64 (15), 1824–1842. <https://doi.org/10.1080/02626667.2019.1678750>.
- Tikhamarine, Y., Malik, A., Pandey, K., Sammen, S.S., Souag-Gamane, D., Heddam, S., Kisi, O., 2020a. Monthly evapotranspiration estimation using optimal climatic parameters: efficacy of hybrid support vector regression integrated with whale optimization algorithm. *Environ. Monit. Assess.* 192, 626. <https://doi.org/10.1007/s10661-020-08659-7>.
- Tikhamarine, Y., Malik, A., Souag-Gamane, D., Kisi, O., 2020b. Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration. *Environ. Sci. Pollut. Res.* 27, 30001–30019. <https://doi.org/10.1007/s11356-020-08792-3>.
- Torres, A.F., Walker, W.R., McKee, M., 2011. Forecasting daily potential evapotranspiration using machine learning and limited climatic data. *Agric Water Manag* 98 (4), 553–562.
- Trajkovic, S., Todorovic, B., Stankovic, M., 2003. Forecasting of reference evapotranspiration by artificial neural networks. *J. Irrig. Drain. Eng.* 129 (6), 454–457.
- Traore, S., Luo, Y., Fipps, G., 2016. Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages. *Agric Water Manag* 163, 363–379.
- Valipour, M., Khoshkam, H., Bateni, S.M., Jun, C., Band, S.S., 2023. Hybrid machine learning and deep learning models for multi-step-ahead daily reference evapotranspiration forecasting in different climate regions across the contiguous United States. *Agric Water Manag* 283, 108311.

- Wen, X., Si, J., He, Z., Wu, J., Shao, H., Yu, H., 2015. Support-vector-machine-based models for modeling daily reference evapotranspiration with limited climatic data in extreme arid regions. *Water Resour. Manag.* 29, 3195–3209.
- Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y., Song, L., 2018. Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *J. Geophys. Res. Atmos.* 123 (16), 8674–8690.
- Yamaç, S.S., Todorovic, M., 2020. Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agric Water Manag* 228, 105875.
- Yu, H., Wen, X., Li, B., Yang, Z., Wu, M., Ma, Y., 2020. Uncertainty analysis of artificial intelligence modeling daily reference evapotranspiration in the northwest end of China. *Comput. Electron. Agric.* 176, 105653.
- Zhao, L., Zhao, X., Zhou, H., Wang, X., Xing, X., 2021. Prediction model for daily reference crop evapotranspiration based on hybrid algorithm and principal components analysis in Southwest China. *Comput. Electron. Agric.* 190, 106424.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat Methodol.)* 67 (2), 301–320.