Smart Agricultural Technology 11 (2025) 100886

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Transfer and deep learning models for daily reference evapotranspiration estimation and forecasting in Spain from local to national scale

Yu Ye, Aurora González-Vidal, Miguel A. Zamora-Izquierdo *, Antonio F. Skarmeta

Department of Information and Communication Engineering, University of Murcia, Murcia, Spain

ARTICLE INFO

Keywords:
Reference evapotranspiration
Transfer learning
Long short-term memory
Generalizability
Spain/Mediterranean climate

ABSTRACT

Accurate estimation and forecasting of Reference Evapotranspiration (ET_0) is critical for almost all agricultural activities and water resource management. However, the most commonly used Penman-Monteith model (FAO56-PM) requires a large amount of input data and it is difficult to compute for general users. Machine Learning (ML) techniques can be used to address this shortcoming. Nevertheless, most studies are site-specific and lack generalizability. This study compares standard ML and Deep Learning (DL) algorithms for estimating and forecasting daily ET_0 at different spatial scales in Spain. While Transfer Learning (TL) is a well-established ML technique, its application in ET_0 computation remains largely unexplored. We applied TL in a novel approach to retrain DL models, enabling adaptation to diverse local climatic conditions, which is particularly important in this domain. All possible combinations of FAO56-PM inputs were evaluated. The results showed that with three or more climatic variables, the TL process can consistently reduce errors by using an appropriate amount of new data to retrain the models. In estimation, with 20% (120 days) of new data, TL models can provide the same performance as if they were trained with local data, both regionally and nationally (improvement of MAE from 26.4% to 99.5%). During forecasting, we used predicted weather data as input, and despite inherent biases in some variables, the TL models successfully adapted using 9-36 days of new data, significantly improving predictive performance (reducing MAE from -1.1% to 134.3%). Thus, the TL process is highly recommended as a promising methodology for increasing the generalization capability of DL models in both daily ET_0 estimation and forecasting under diverse climatic conditions with limited local data.

1. Introduction

Water is one of the most important resources for human beings and according to the Food and Agriculture Organization (FAO), agriculture is the most water-scarce sector. Currently, more than 90% of freshwater consumption across the world is attributed to agriculture, which also accounts for 70% of freshwater withdrawals globally. Furthermore, the FAO warns that climate change may significantly exacerbate water scarcity [14].

In agricultural contexts, evapotranspiration (ET) is the principal component of the hydrological cycle and represents the combined loss of water from two primary sources: evaporation from the soil surface and transpiration from plants. Evaporation and transpiration occur simultaneously and are affected by weather parameters, crop characteristics, management and environmental factors, and there is no easy way to distinguish between the two processes [27].

The crop evapotranspiration under standard conditions (ET_c) represents the actual water needs of a specific crop under given environmental conditions, taking into account the crop characteristics and growth stage [35]. ET_c can be obtained by multiplying the reference crop evapotranspiration (ET_0), which is a standardized measure based on the potential water loss from a well-watered reference area, by the specific crop coefficient (K_c) [5].

Accurately calculating ET_0 is therefore crucial for managing water resources and scheduling irrigation, especially in arid and semi-arid areas such as Spain and most of the Mediterranean countries. It can be directly measured by a lysimeter or estimated by calculating different physical-mathematical models [27,18,35]. When no physical instrument is available, the FAO56-Penman-Monteith (FAO56-PM) equation is the most widely accepted standard equation for ET_0 measurement among all estimation methods and has been validated worldwide under different climatic conditions [4,42,51,46]. Nevertheless, the main constraint of this method is that it requires numerous inputs, mainly meteorolog-

* Corresponding author.
E-mail address: mzamora@um.es (M.A. Zamora-Izquierdo).

https://doi.org/10.1016/j.atech.2025.100886
Received 27 December 2024; Received in revised form 6 March 2025; Accepted 6 March 2025
Available online 12 March 2025

ical variables, i.e., air temperature (T), relative humidity (RH), wind speed at 2 m height (U_2) and solar radiation (Rs), which are generally not available at a given location, especially in developing countries [6]. In addition, it also requires other specific variables such as latitude, day of the year and the height of the wind speed sensor (if it is measured different from two meters above the ground).

In recent decades, Artificial Intelligence (AI) and Machine Learning (ML) have been utilized in various scientific fields, including the estimation and forecasting of ET_0. To address problems with non-linear relationships between variables (inputs and target), the use of AI models is an efficient way to identify and select inputs, as well as to discover the appropriate relationship between variables [43]. Therefore, the use of ML is suitable in scenarios where not all meteorological variables are available for calculating ET_0, enabling the generalization of a complex model such as the FAO56-PM, which is still the reference model. Furthermore, one of the main advantages of using these ML models is that they are a feasible alternative to the FAO56-PM equation even with all available weather parameters, since they eliminate the need to calculate derived parameters that require specific domain knowledge by directly calculating ET_0 values, making them useful for irrigation managers and practitioners who may lack technical knowledge [20].

Some of the most commonly used ML models are ensemble learning models such as Random Forest (RF) or Extreme Gradient (XG)-Boost [15,29,23,37]; Support Vector Machine (SVM) [23,17,41,26,22]; and different implementations of Artificial Neural Networks (ANN) such as Multi Layer Perceptron (MLP), Generalized Regression Neural Network (GRNN) or Extreme Learning Machine (ELM) [15,37,9,17,41,26].

On the one hand, for estimation, Huang et al. [23] used CatBoost, RF and SVM with different combinations of climatological variables as inputs to estimate ET_0 in humid regions of China. They found that SVM offered the best predicting accuracy and stability with incomplete input variables and CatBoost performed best when all meteorological variables are used. Bellido-Jiménez et al. [8] estimated daily ET_0 using only temperature-based data in southern Spain and compared the performance of different temperature-based methods (HS and FAO56-PMT) with a variety of ML models (MLP, ELM, GRNN, SVM, RF, and XGBoost). They found that ELM and MLP had better performance among all stations used with $R^2 > 0.89$, NSE > 0.88, RMSE < 0.67 mm d^{-1} and MBE between -0.17 and 0.3 mm d^{-1}. Ferreira et al. [17] grouped 203 weather stations with similar climatic characteristics (temperature and relative humidity) using K-means, and then evaluated both ML models and empirical equations for estimating ET_0 in Brazil, being ANNs the best evaluated alternatives.

On the other hand, for ET_0 forecasting, there could be two main methods: direct and indirect [44]. The first approach uses historical data in order to train models [32,2,16,21,30] while the second is to forecast the necessary weather parameters for the ET_0, most of which can be achieved using Numerical Weather Predictions (NWP) [49,13,45,31]. The latter method can give good results, but the forecasting performance of the meteorological data is crucial. Perera et al. [36] used NWP forecasts derived from the Australian Community Climate and Earth System Simulator-Global (ACCESS-G) to forecast daily ET_0 for lead times up to 9 days (using FAO56-PM) at 40 geographically diverse locations across Australia. They found that daily ET_0 forecasts calculated using NWP weather variable forecasts were better than the long-term monthly mean ET_0 for lead times up to 6 days. They also concluded that errors in incoming solar radiation were the most important source of ET_0 forecast error and, together with wind speed, showed the worst forecast performance. Liu et al. [28] compared the ET_0 obtained from the output of both NWP and PWF (Public Weather Forecast) with the real data at 31 stations across China. NWP had better forecast performance in the mountain plateau and temperate continental zones (R = 0.84, MAE = 0.63 mm d^{-1}, RMSE = 0.86 mm d^{-1} and R = 0.86, MAE = 1.32 mm d^{-1}, RMSE = 1.94 mm d^{-1}, respectively), and PWF can provide a more accurate ET_0 forecast in the temperate monsoon and subtropical monsoon zones (R = 0.89, MAE = 1.00 mm d^{-1}, RMSE = 1.39 mm d^{-1} and R = 0.80, MAE = 1.36 mm d^{-1}, RMSE = 1.68 mm d^{-1}, respectively).

The use of AI in the area of ET_0 is well researched. However, based on the literature review, most of the ET_0 studies using AI have focused on building models at a specific site, therefore the main drawback of this type of methodology is the limitation of significant performance variation, i.e., they lack generalizability [3]. In order to address the issue, this study proposes an approach based on Transfer Learning (TL), which enables DL models to adapt to new climate conditions using pre-trained models and adapting them to the new data. The major contributions of this work are listed below:

- A novel approach using the TL technique is employed to improve the adaptability of DL models (LSTM) across different local levels.
- An analysis of all combinations of input weather variables used by FAO56-PM (i.e. T, RH, U_2, and Rs) is provided.
- The validation of models at local, regional, and national scales for both historical estimates and future forecasts of ET_0 by comparing standard ML models (RF, SVR, and MLP) with TL models.

The remainder of this paper is organized as follows: Section 2 presents the materials and methods used for this study. Then, the detailed results archived at different scales are provided in Section 3. Section 4 discusses the results and finally, Section 5 concludes the paper and highlights the directions for future research.

2. Material and methods

2.1. Study area and data acquisition

The present study was carried out in Spain, focusing on the region of Murcia as a reference for the creation of the pre-trained models. Murcia is located in the southeastern part of the Iberian Peninsula, on the Mediterranean coast. The region covers an area of 11,313 km^2 and has a diverse topography that includes plains, mountains and coastal areas. The climate is arid and semi-arid, with hot summers and mild winters. In terms of water resources, the region has limited rainfall (less than 350 mm/year) and drought intensely marks landscapes, agricultural activities and rural and urban concerns. The southeastern part of Spain, where the Region of Murcia is located, is one of the driest areas in Europe.

On a national scale, only about 40% of Spain's land is suitable for cultivation due to its varied topography, with the main agricultural areas in the northern and Central Plateaux. The Mediterranean climate brings hot, dry summers and cold, wet winters. Precipitation, concentrated between October and April (south) or May (north), is characterized by considerable inter-annual variability. Summer rainfall is generally less than 100 mm, except in the northwest, requiring additional irrigation for crop production. The dry season has a coefficient of variation of precipitation ranging from 21 to 55%, posing a high risk of failure for rainfed crops. Supplemental irrigation is essential to meet water needs and achieve potential yields [24].

There are three groups of Automatic Weather Stations (AWS) where real data is collected, as shown in Fig. 1. The first group has three stations from the Agricultural Information System of Murcia (SIAM), which will be used to train the models and validate them. The second group has five AWS (also from SIAM) to test both the estimation and the forecast (same as the third group). The data for the last group are obtained from the Spanish Agroclimatic Information System for Irrigation (SIAR), with a total of 27 stations, covering the main Köppen climate classifications of Spain [7] and taking into account the different environmental conditions and morphological characteristics of the peninsula. All data records from both sources, SIAM and SIAR, are openly available and can be downloaded at http://siam.imida.es/apex/f?p=101:46:4060071267268956 and https://servicio.mapa.gob.es/websiar/SeleccionParametrosMap.aspx?dst=1, respectively. Table 1 shows the

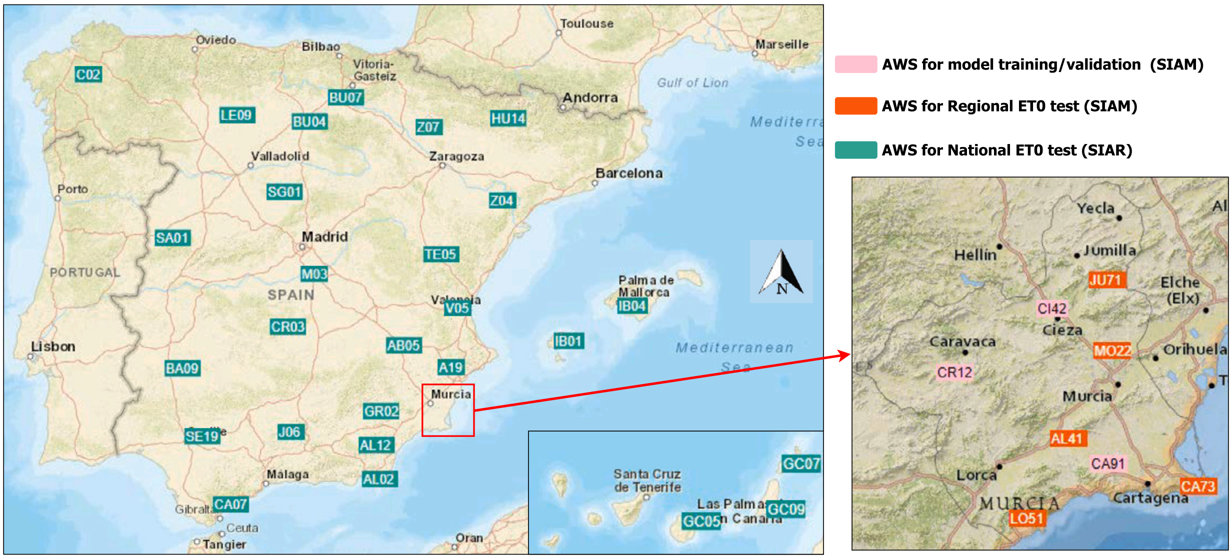


Fig. 1. Map of all AWS used in Spain and the Region of Murcia.

geographical information and mean daily values of meteorological variables selected from three groups of stations.

A Python script is implemented to download the daily forecast of the climatic variables from WeatherBit [47] and Open-Meteo [33] for the forecast set period. Regional forecasts are obtained from WeatherBit, which provides weather forecasts that are composites of many of the world's major NWP models, including the European Centre for Medium-Range Weather Forecasts (ECMWF), Global Forecast System (GFS) and Icosahedral Nonhydrostatic (ICON). These models are used together with real-time observations and post-processing to produce a worldwide forecast. Open-Meteo, on the other hand, is an open-source API and provides the worlds most used weather models. In this study, NOAA GFS model is used to collect national forecasts.

2.2. FAO56-Penman-Monteith

The target value used in this study is calculated from the FAO56-PM equation [5]:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \tag{1}$$

where ET_0 is the reference evapotranspiration (mm day^{-1}), R_n is the net radiation at the crop surface (MJ m^{-2} day^{-1}), G is the soil heat flux density (MJ m^{-2} day^{-1}) which is assumed to be zero when the calculation is for daily values, T is the mean daily air temperature (°C) calculated as the mean of Tmax and Tmin, U_2 is the wind speed at 2 m height (m s^{-1}), e_s is the saturation vapor pressure at the mean daily air temperature (kPa), e_a is the actual vapor pressure (kPa), Δ is the slope of the saturation vapor pressure curve (kPa °C^{-1}), and γ is the psychrometric constant (kPa °C^{-1}).

All wind speed data monitored at a different height than the 2 m standard was converted to such a standard as follows:

$$U_2 = u_z\frac{4.87}{\ln(67.8z - 5.42)} \tag{2}$$

where u_z is the measured wind speed at z m above the ground surface (m s^{-1}).

2.3. Machine learning (ML) algorithms

The selected ML algorithms are RF, SVR, and MLP since they are capable of capturing and modeling complex nonlinear relationships in the data and have been widely used in a variety of hydrological applications, including ET_0 estimation [20]. The main reason for not using time series models (such as ARIMA or SARIMA) is that ET_0 is a complex and dynamic hydrological process that depends on various meteorological variables and the interactions between them [20], and these models are mainly focused on pattern recognition and forecasting rather than modeling the relationship between inputs and outputs. Moreover, missing values could be a major challenge for these time series models as well [38,19].

2.3.1. Random forest (RF)

RF is a powerful ML algorithm in the ensemble learning category [10]. RF relies on decision trees trained through the Classification and Regression Tree (CART) algorithm, which allows decisions to be made based on input characteristics. The core process involves creating an ensemble, or collection, of decision trees, each generated using random subsampling at two levels: feature selection and data sampling. At each tree split, only a random subset of features is considered, which promotes diversity among trees and prevents high correlation between them. Additionally, each tree is trained on a random, bootstrapped subset of the data, adding variability and enhancing the model's ability to handle noise and avoid overfitting.

For regression predictions, RF uses averages of the outputs of the trees:

$$RF(X) = \frac{1}{N}\sum_{i=1}^{N} T_i(X) \tag{3}$$

where X is the input features and $T_i(X)$ the prediction of the i-th decision tree.

2.3.2. Support vector machine (SVM)

SVM is a robust ML algorithm that seeks to find a hyperplane with the maximum margin between classes [12]. In the linear case, it aims to minimize the norm of the weight vector while ensuring that all data points are correctly classified. When data is not perfectly separable, soft margin SVM introduces a slack variable, allowing some misclassifications and balancing margin width with classification accuracy through a regularization parameter, C.

For regression tasks, Support Vector Regression (SVR) adapts SVM to fit a hyperplane within a specified error margin. This objective function includes a regularization term and an insensitive loss term to handle errors within this margin.

Table 1
Geographical information and mean daily meteorological variables values of all stations used.

Type	Code	Location	Lat. & Lon.	Alt. (m)	ET_0 (mm d^{-1})	T_{max} (ºC)	T_{min} (ºC)	RH_{max} (%)	RH_{min} (%)	R_s (w/m^2)	U_2 (ms^{-1})
Local/Regional	CA91	Fuente Álamo	37.6990,-1.2380	175	3.21	23.6	12.38	83.25	38.24	210.94	0.95
Local/Regional	CI42	Cieza	38.2838,-1.4963	244	3.48	24.36	10.92	86.74	35.79	208.69	1.3
Local/Regional	CR12	Caravaca	38.0439,-1.9800	869	3.36	19.86	7.54	80.11	36.59	201.83	1.85
Regional	AL41	Alhama	37.7922,-1.4275	169	3.97	24.25	10.98	81.1	34.66	210.27	1.98
Regional	CA73	Cartagena	37.6111,-0.8037	92	3.03	21.94	15.55	87.16	54.27	191.23	1.46
Regional	JU71	Jumilla	38.3944,-1.2393	401	3.7	23.51	10.11	81.15	33.03	208.71	1.74
Regional	LO51	Aguilas	37.4879,-1.6239	329	3.83	22.35	14.68	78.63	41.21	211.07	2.11
Regional	MO22	Molina Segura	38.1275,-1.2206	146	3.84	24.48	12.86	79.75	32.24	206.19	1.69
National	A19	Villena	38.5955, -0.875	488	3.56	22.56	8.29	91.03	39.67	201.55	2.07
National	AB05	Albacete	38.9493,-1.9017	689	3.56	21.55	7.07	92.25	40.33	205.33	2.22
National	AL02	Almería	36.8354,-2.4024	5	3.39	23.68	14.52	90.42	46.59	213.49	1.25
National	AL12	Tíjola	37.3785,-2.4596	776	3.41	21.67	9.79	83.71	35.67	206.39	1.56
National	BA09	V. de los Barros	38.5755,-6.3485	403	3.58	23.59	10.63	82.26	37.45	200.32	1.76
National	BU04	Tardajos	42.3461,-3.8026	824	2.81	17.81	5.17	94.78	48.51	178.39	2.17
National	BU07	Santa Gadea del Cid	42.7018, -3.075	525	2.6	19.18	6.17	95.09	49.75	159.56	1.71
National	C02	Boimorto	43.0338,-8.1417	412	2.19	17.69	7.74	96.09	60.56	148.32	1.71
National	CA07	J. de la Frontera	36.4136,-5.3838	50	3.21	24.24	11.59	90.35	43.88	205.18	1.19
National	CR03	Porzuna	39.2343,-4.2307	610	3.29	22.54	7.92	88.08	37.95	201.12	1.52
National	GC05	Arucas	28.1294,-15.514	220	3.09	21.9	16.86	88.17	59.06	184.6	1.74
National	GC07	Tinajo	29.0500,-13.659	254	3.64	22.8	15.63	91.63	55.68	196.91	4.02
National	GC09	Antigua - Pozo Negro	28.3325, -13.94	68	4.65	24.87	17.47	78.15	44.9	224.19	3.16
National	GR02	P. de Don Fadrique	37.8761, -2.381	1017	3.78	20.9	6.12	88.89	33.82	209.99	2.58
National	HU14	Banastón	42.3903,0.1694	559	2.59	21.41	6.23	95.24	40.64	179.1	0.76
National	IB01	Santa Eulalia	39.0100,1.43992	120	2.98	23.6	10.98	96.74	47.48	183.74	1.28
National	IB04	Son Ferriol	39.5623,2.72610	8	2.88	23.76	11.38	93.48	47.78	183.61	1.01
National	J06	Alcaudete	37.5772,-4.0783	640	3.09	23.2	11.46	77.98	35.88	202.91	0.86
National	LE09	Santas Martas	42.4410,-5.2601	876	3.16	18.25	5.76	91.99	44.32	189.08	2.59
National	M03	Aranjuez	40.0416,-3.6304	486	3	23.41	7.34	90.32	36.33	195.7	1.05
National	SA01	Ciudad Rodrigo	40.5901,-6.5386	616	2.76	20.72	5.99	95.13	44.28	188.05	1.17
National	SE19	IFAPA C. Torres-Tomejil	37.5125,-5.9641	12	3.21	26.23	11.59	90.89	36.69	204.39	0.84
National	SG01	Gomezserracín	41.3014, -4.298	813	3.03	19.75	3.72	94.65	41.54	188.95	1.78
National	TE05	Teruel	40.3472, -1.166	913	2.75	20.62	5.05	93.14	35.8	186.44	1.03
National	V05	Cheste	39.5188,-0.7444	315	2.76	23.59	9.57	92.47	41.43	184.43	0.77
National	Z04	Fabara	41.1678,0.15403	245	3.56	22.62	9.23	92.29	41.99	191.67	2.25
National	Z07	Sádaba	42.2673,-1.3093	432	3.26	20.51	7.98	92.51	45.98	182.03	2.51

In non-linear scenarios, SVM employs different kernel functions to handle complex decision boundaries. The popular Radial Basis Function (RBF) kernel maps data into a higher-dimensional space, allowing the SVM to classify data that is not linearly separable in the original feature space.

2.3.3. Multi layer perceptron (MLP)

An ANN is a computational model inspired by biological neural networks. Its basic unit, the neuron, processes inputs via an activation function to produce an output. Neurons are organized into layers: the input layer receives data, hidden layers handle intermediate processing, and the output layer provides the final result. The output of each neuron is calculated by applying an activation function to a weighted sum of inputs plus a bias term:

$$a_j = f(\sum_{i=1}^{n} w_{ij} x_i + b_j) \tag{4}$$

where a_j is the output of node j, $f(\cdot)$ is the activation function, w_{ij} is the weight connecting node i in the previous layer to node j, x_i is the input to node i, and b_j is the bias term for node j.

Activation functions such as Sigmoid, Tanh, and ReLU introduce non-linearity, allowing ANNs to learn complex patterns. A common ANN structure, the MLP, consists of multiple layers, with information flowing forward through each layer.

Training involves minimizing a loss function to reduce the difference between predicted and actual values. This is achieved by backpropagation, with weights and biases adjusted by optimizers such as Adam, SGD, or RMSprop. Key training parameters include learning rate, number of hidden layers and neurons, choice of activation function, batch size, and number of epochs.

2.4. Transfer learning and deep learning

Transfer learning is a ML technique where a model developed for one task is reused as the starting point for a model on a second, related task. Instead of training a model from scratch, TL leverages the patterns and features learned from a large dataset (usually for a similar task) to improve performance and speed up training on a smaller, more specific dataset.

In deep learning, TL refers to the practice of reusing the weights in one or more layers of a pre-trained neural network, often one trained on a large dataset, as the starting point for a new model. The retraining process can either keep the weights of some layers fixed (usually the lower layers) and update the weights of the remaining layers (usually the upper layers), or completely adapt all the weights of the neural network [25]. Fixing (or freezing) a layer's weight can preserve the features learned by the original model, and it can avoid overfitting during the retraining process. When all hidden layers of a neural network are fixed and only the output layer is updated, the TL process is used as a feature extraction method. Conversely, changing (or fine-tuning) the original weights allows the model to apply past knowledge in the target domain and re-learn some new insights.

The TL process used in this study is shown in Fig. 2. The S_1 corresponds to the three training stations and the S_2 refers to the remaining stations. Therefore, the original models are trained from the training stations and then retrained with the target station to obtain the TL model. Both Fine-Tuning (FT) and Freezing all layers (Fr) methods, along with the percentage of new data to retrain the model, are tested as parameters for the TL process.

The DL model used in this work is Long-Short Term Memory (LSTM), a type of Recurrent Neural Network (RNN) designed to process and predict time-series data. LSTM was introduced in 1997 [39] and it addresses

Table 2
RF, SVR and MLP hyperparameter searchspaces.

Algorithm	Hyperparameter	Search space
RF	Maximum depth of the tree	From 2 to 200 with 2 intervals
	Number of trees in the forest	From 10 to 2000 with 10 intervals
	Measure of the quality of a split	squared_error, absolute_error, friedman_mse and poisson
	Number of features to consider when looking for the best split	sqrt, log2 and None
SVR	Kernel	RBF
	Regularization parameter	0.01, 0.1, 1, 10, 50, 100 and 1000
	Kernel coefficient	0.1, 1, 10, 20 and 50
	Epsilon	0.01, 0.1, 1 and 10
MLP	Neurons in the 1st hidden layer	From 12 to 1024 with 12 intervals
	Neurons in the 2nd hidden layer	From 4 to 256 with 4 intervals
	Dropout between hidden layers	0, 0.2, 0.4, 0.6 and 0.8
	Loss functions	MSE and MAE
	Optimizer	Adam, SGD, and RMSprop
	Learning rate	0.0001, 0.001, 0.01 and 0.1
	Activation	ReLU and Tanh
	Batch size	32, 64 and 128
LSTM	The same as MLP	The same as MLP

the problem of long-term dependency that is common in standard RNNs. LSTM has special units called memory cells that can store information for long periods of time. This allows them to retain and use information from earlier in a sequence, which is crucial for tasks involving time series or sequential data.

LSTM uses three types of gates to regulate the flow of information: input gate (i_t), forget gate (f_t), and output gate (o_t), which can be expressed as follows:

$$f_t = \sigma(W_f \cdot h_{t-1} + W_f \cdot x_t + b_f), \tag{5}$$

$$i_t = \sigma(W_{ih} \cdot h_{t-1} + W_{ix} \cdot x_t + b_i), \tag{6}$$

$$\tilde{c}_t = \tanh(W_{\tilde{c}h} \cdot h_{t-1} + W_{\tilde{c}x} \cdot x_t + b_{\tilde{c}}), \tag{7}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t, \tag{8}$$

$$o_t = \sigma(W_{oh} \cdot h_{t-1} + W_{ox} \cdot x_t + b_o), \tag{9}$$

$$h_t = o_t \cdot \tanh(c_t), \tag{10}$$

where σ and $\tanh$ are activation functions; W_f and b_f are the weights and bias for f_t; h_{t-1} is the previous hidden state, and x_t is the current input. The cell state c_t is updated by combining the old cell state and the new candidate values $\tilde{c}_t$. o_t controls which parts of the c_t contribute to the hidden state h_t that is passed to the next time step. Finally, h_t is updated.

2.5. Experiment configurations and study workflow

All simulations in this study were performed on a computer with an AMD Ryzen 7 3700X 8-core processor at 3.59 GHz and 16 GB of RAM. All ML models and their training process were implemented using the following Python libraries: scikit-learn [34], Keras [11], SciKeras [40] and TensorFlow [1]. In order to ensure the reproducibility of the results, all processes with randomization options are configured with a seed of the value 123. For reproducibility purposes, all the functions and scripts needed to run the experiments, along with the collected data, are available at https://github.com/yuye188/ML_ET0_Spain.

The complete workflow is shown in Fig. 3. In this study, we refer to Tmax, Tmin as **T** and RHmax, RHmin as **RH**, since these values are obtained simultaneously from the same source. Mean temperature and relative humidity are not used because the FAO56-PM uses the maximum and minimum of these variables. Therefore, four climatic variables are considered (T, RH, U, Rs), which are available in almost all the Autonomous Communities of Spain.

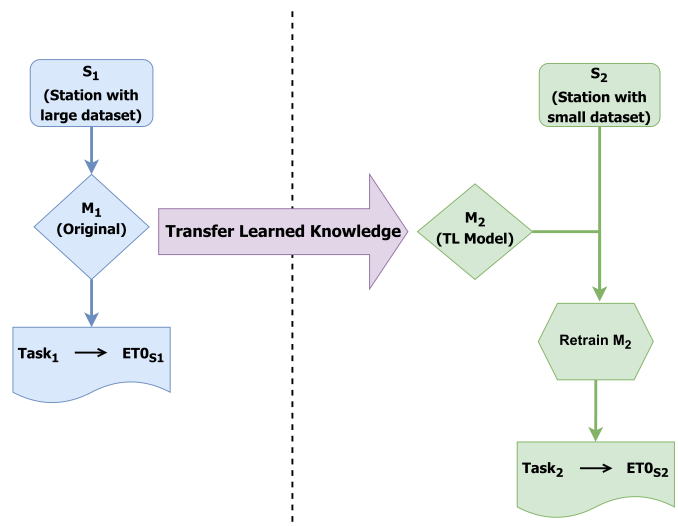


Fig. 2. TL process.

During the ML training phase, the hyperparameter search is performed with randomizedSearchCV with 5-fold Cross-Validation (CV5) and 60 iterations, taking advantage of parallel processing capabilities.

The hyperparameters tuned for each algorithm are listed in Table 2. In MLP and LSTM, in addition to the shown options, an early stop with patience 30 (within a total of 500 epochs) is set to avoid overfitting and to stop training as soon as performance on a validation set stops improving.

To build the models, data from the training stations are divided into training and validation sets. For each station, combinations of four algorithms with four climatic variables are trained with hyperparameter tuning and then evaluated with the validation set. In this way, with three training stations and four algorithms, the best models are obtained with four, three, two, and one variables (denoted from M4 to M1), for a total of 48 models. To simplify the comparison between models without TL (RF, SVR and MLP) and with TL (LSTM), we selected the best 12 models (four in three stations for each number of variables) from the former list. The MAE validation is used to select the best model from four combinations of variables as it is easy to interpret, provides a simple but direct measure of the average prediction error, and is less sensitive to outliers than other metrics such as RMSE or MAPE. Finally, there are 12 models for both TL and No-TL models to compare.

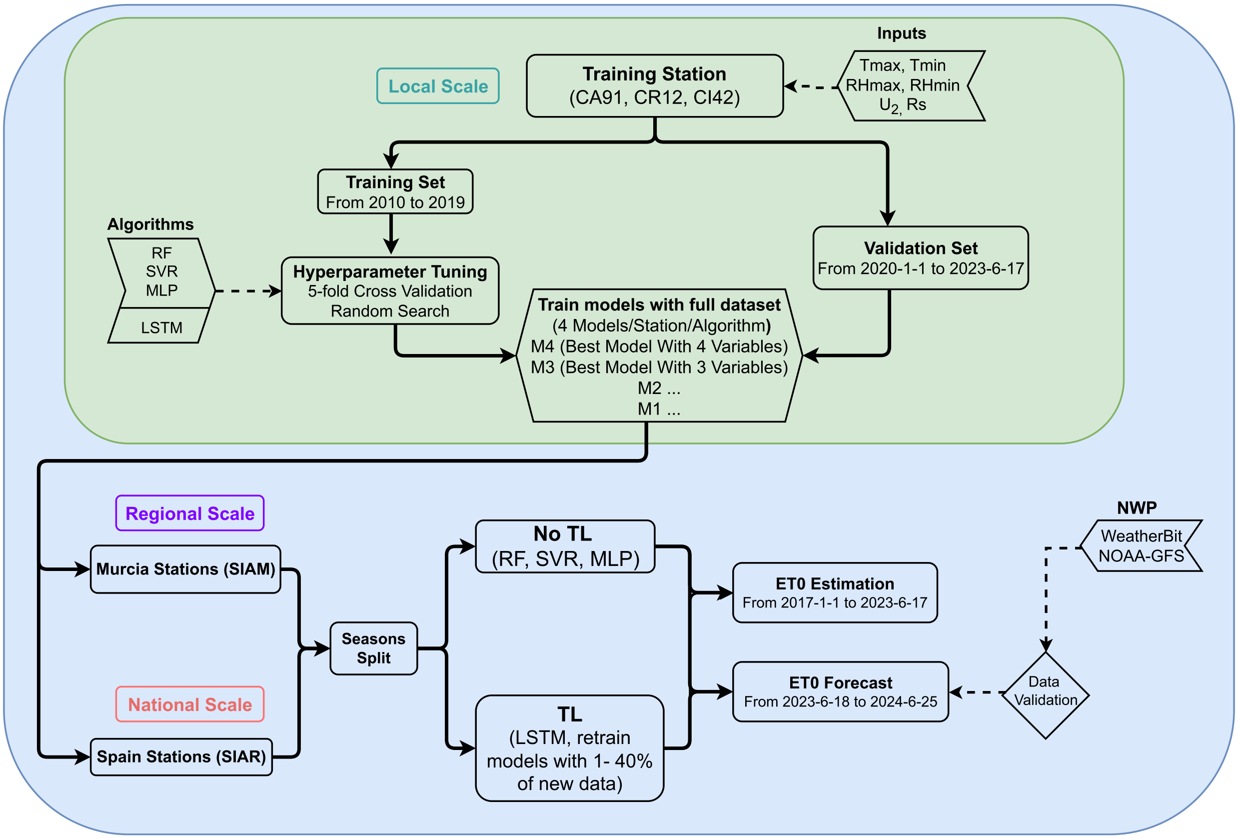


Fig. 3. The workflow of this study.

To test the models, data from regional and national stations are divided into an estimation set and a forecast set. Before evaluation, both sets are divided into four seasons. The reason for this step is that the pattern of ET_0 is strongly influenced by seasonal factors, then if the data used in the retraining process covers only one season, the performance of the models will not improve for the other three seasons. Moreover, in a real use case, having a pre-trained model and then fine-tuning it on recent seasonal data, the model can use the general features initially learned, but become more specialized through seasonal fine-tuning. This helps to efficiently adapt the model to new conditions each season. After seasons split process, different percentage of each season will be used to retrain the TL models. For estimation tests where the datasets contain more observations, 1, 5, 10, and 20% of the initial data are used to retrain the TL models, corresponding to 6, 30, 60, and 120 days. On the other hand, for one-year data, 10, 20, 30, and 40% of the initial data (9, 18, 27, and 36 days, respectively) are used for forecasting tests. The data used to retrain the TL models are removed when testing the NO-TL models to ensure that both methods are evaluating the same test sets. Finally, days with missing data are also removed from all datasets.

2.6. Evaluation metrics

Coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used to evaluate model performance, which are expressed as follows:

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(Y_i - \bar{Y}_i)\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - Y_i| \tag{12}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - Y_i}{y_i}\right| \times 100 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - Y_i)^2} \tag{14}$$

since MAPE is very sensitive to outliers, the Coefficient of Variation of the Root Mean Square Error (CVRMSE) is used in some cases:

$$CVRMSE = (\frac{RMSE}{\bar{y}}) \times 100 \tag{15}$$

where n is the number of data points, y_i and Y_i present the observed (real) and predicted value of the ET_0 for the i-th data point (day); $\bar{y}$ and $\bar{Y}$ are the mean of the observed and predicted values of y and Y.

As an important requirement for ML algorithms, all input variables are standardized using the StandardScaler implementation of the Sklearn library, which follows:

$$z = \frac{\mathrm{x} - \mathrm{u}}{\mathrm{s}} \tag{16}$$

where z is the standardized value, x is the original value, u is the mean of the feature and s is the standard deviation of the feature.

Finally, in the discussion section, we will also use the percentage of MAE improvement comparing the models with and without TL process and it is expressed as:

$$MAE_{improvement} = \frac{MAE_{No-TL} - MAE_{TL}}{MAE_{TL}} * 100 \tag{17}$$

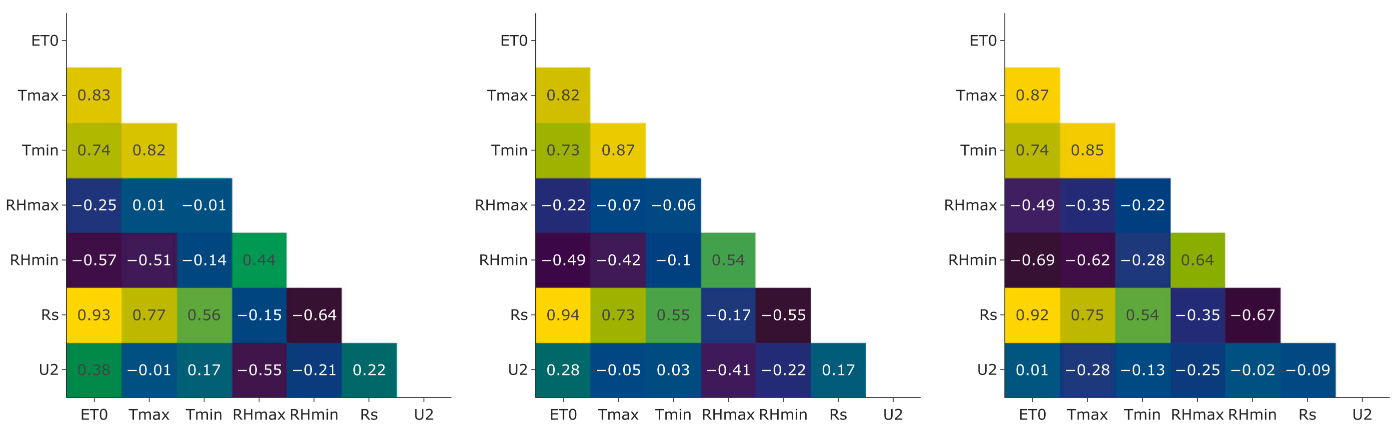
3. Results

This section gathers the results of this study organized into three parts: i) Validation of the models at local scale; ii) ET_0 estimation and forecasting at regional scale; iii) ET_0 estimation and forecasting at national scale. The comparison of TL and No-TL models is done by comparing the best or best average result achieved by each side, the full results at each scale are presented in Appendix A.

Table 3
Statistical performance of local models in the validation set with four best combinations of climate variables at three training stations. The bold values indicate the best statistical indicators obtained within the four algorithms.

Station	Combinations	R2				MAE (mm d^{-1})				MAPE (%)				RMSE (mm d^{-1})			
		RF	SVR	MLP	LSTM	RF	SVR	MLP	LSTM	RF	SVR	MLP	LSTM	RF	SVR	MLP	LSTM
CI42	T, RH, Rs, U2	0.986	**0.990**	0.989	0.989	0.174	**0.147**	0.150	0.149	8.989	7.388	7.593	**7.360**	0.229	0.208	0.210	**0.205**
	T, Rs, U2	0.986	0.989	**0.990**	0.990	0.180	0.154	0.154	**0.147**	8.568	7.659	7.434	**7.320**	0.236	0.212	**0.202**	0.205
	T, Rs	0.972	**0.976**	0.976	0.974	0.370	0.337	0.400	**0.326**	19.243	16.904	18.521	**15.963**	0.447	0.411	0.477	**0.402**
	Rs	0.905	**0.905**	0.904	0.904	0.549	**0.546**	0.547	0.568	29.625	27.644	**25.230**	33.432	**0.689**	0.696	0.704	0.698
CA91	T, RH, Rs, U2	0.984	0.986	**0.987**	0.987	0.158	0.142	**0.138**	0.142	7.517	6.710	**6.506**	6.658	0.218	0.212	**0.208**	0.218
	T, Rs, U2	0.984	**0.986**	0.986	0.985	0.154	**0.142**	0.144	0.149	7.675	**6.941**	7.043	7.230	0.217	**0.209**	0.216	0.225
	T, Rs	0.970	**0.974**	0.972	0.972	0.255	0.220	0.232	**0.208**	13.188	11.036	12.549	**10.352**	0.325	**0.291**	0.304	0.292
	Rs	0.907	**0.908**	0.907	0.906	0.432	**0.420**	0.433	0.452	20.289	18.391	19.485	**17.586**	**0.540**	0.548	0.553	0.573
CR12	T, RH, Rs, U2	0.984	0.990	0.990	**0.991**	0.171	0.122	0.124	**0.119**	7.651	5.908	5.845	**5.420**	0.224	0.172	0.177	**0.166**
	T, Rs, U2	0.982	0.985	0.984	**0.986**	0.176	**0.152**	0.154	0.154	8.385	7.373	7.636	**7.364**	0.231	**0.212**	0.219	0.213
	T, Rs	0.948	0.951	0.952	**0.953**	0.363	0.329	**0.286**	0.369	16.117	14.009	**12.630**	14.152	0.458	0.430	**0.389**	0.469
	Rs	0.858	0.859	0.858	**0.860**	0.530	**0.513**	0.534	0.551	23.965	**21.791**	23.939	22.835	0.679	**0.673**	0.682	0.716

Note: T and RH include the max and min of each variable.



(a) Station CI42 (b) Station CA91 (c) Station CR12

Fig. 4. Correlation plot between variables in three training stations (CI42, CA91, CR12).

3.1. Validation of the models at local scale

There are many combinations of variables that can be done for inputs. As already explained, we have tried all combinations of variables with one, two, three, and four inputs, respectively, and the best one for each of them is shown in Table 3.

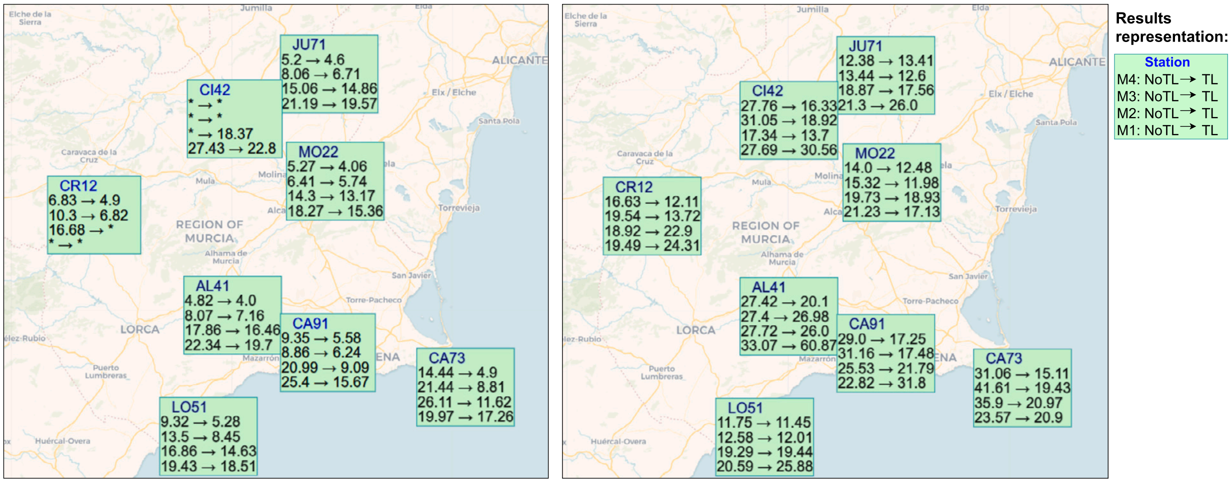
For all three stations, the best combination is the same for each number of variables. For M1 models, Rs is undoubtedly the most useful variable for estimating ET_0; therefore, it is used as the basis for all four best combinations. T and Rs are the best with two variables (M2). Then T, Rs, and U_2 as M3, and finally M4 with all variables. Considering those results, the variable importance can be ordered as $Rs > T > U_2 > RH$, similar with the correlation of the variables shown in Fig. 4, where Rs always has a correlation greater than 0.9 with ET_0. The minimum of T and RH are more correlated than their maximum values, and despite the almost inexistent correlation between U_2 and ET_0 in CR12, it seems that U_2 does have a significant contribution to estimate ET_0 for all models.

At station CI42, the full variable combination produced the highest performance metrics. SVR achieved the best R^2 value of 0.99, and LSTM demonstrated the lowest MAE (0.147 mm d^{-1}), suggesting it minimized errors effectively. However, the combination for M3 models also performed well, with MLP reaching an R^2 of 0.99 and showing a competitive RMSE of 0.205 mm d^{-1}, indicating that dropping RH had minimal impact on these models at CI42. When using fewer variables, such as T and Rs, SVR's R^2 decreased slightly to 0.976, and LSTM's RMSE increased to 0.402 mm d^{-1}. Using only Rs as input led to significantly lower performance, with the highest R^2 being 0.905 (SVR) and the lowest RMSE being 0.689 mm d^{-1} (RF). This illustrates the benefits of including multiple variables for enhanced accuracy, as the reduced variable sets led to higher errors.

For station CA91, the pattern was similar. Here, MLP achieved the highest R^2 of 0.987, and the lowest MAE (0.138 mm d^{-1}), demonstrating strong predictive power with all variables included. With the T, Rs, U2 combination, SVR still performed well, achieving an R^2 of 0.986, while MLP maintained a low RMSE of 0.208 d^{-1}. Reducing the variables to T, Rs led to a drop in performance, with SVR's R^2 decreasing to 0.974 and RMSE value increased to 0.291 mm d^{-1}. When only Rs was used, performance was notably lower, with SVR achieving an R^2 of 0.908 and RF yielding the lowest RMSE at 0.54 mm d^{-1}.

In station CR12, the best performance was again observed with the full variables combination, where LSTM achieved the highest R^2 of 0.991, lowest MAE and MAPE (0.119 mm d^{-1} and 5.42%, respectively). For the M3 combination, SVR remained competitive, with an MAE of 0.152 mm d^{-1} and an RMSE of 0.212 mm d^{-1}. With fewer variables in the T and Rs combination, LSTM's R^2 dropped to 0.953 and MAPE increased to 7.364%, but these values were still reasonably accurate. Finally, using only Rs led to the lowest performance, as SVR's MAE dropped further to 0.513 mm d^{-1} and MAPE to 21.791%.

As it can be seen, in most cases, adding a climatic variable provides a large improvement, reducing MAE, MAPE and RMSE by 30-50% in all models, except for RH. However, in CR12, although located in an arid region, due to its weather characteristics, which are more like a humid climate, with the lowest average temperature and radiation, and also more windy (see Table 1), there is a noticeable improvement when using RH. Therefore, the importance of RH cannot be neglected. However, de-



(a) Murcia ET_0 estimation improvements using TL (b) Murcia ET_0 forecast improvements using TL

Fig. 5. Comparison of regional ET_0 estimation (left) and forecasting (right) results between the best performing No-TL and TL models. The green box represents the results (MAPE) in the corresponding station. The stations CA91, CI42 and CR12 were used as train stations, therefore they cannot be included as a test in the estimation of their own station.

Table 4
Best mean regional ET_0 estimation statistical metrics using 12 No-TL and 12 TL models trained from three stations (CI42, CA91, and CR12). The bold values indicate the best values obtained between No-TL and TL models, with the corresponding TLPercentage used to retrain the TL models. The suffix FT in TL models indicates that the model uses the fine-tuning method and Fr indicates that the model uses the freezing all layers method.

Models	TLPercentage	R^2		MAE (mm d^{-1})		MAPE (%)		RMSE (mm d^{-1})	
		NO-TL	TL	NO-TL	TL	NO-TL	TL	NO-TL	TL
M4	1% (6 days)	0.962	0.965	0.178	0.225	7.721	9.083	0.239	0.276
NO-TL:CI42-SVR	5% (30 days)	0.962	0.969	0.178	0.179	7.729	7.371	0.239	0.226
TL: CI42-LSTM-FT	10% (60 days)	0.962	0.975	0.179	0.145	7.790	6.115	0.241	0.186
	20% (120 days)	0.962	**0.978**	0.179	**0.116**	7.889	**4.761**	0.238	**0.153**
M3	1% (6 days)	0.926	0.931	0.253	0.253	10.643	10.730	0.330	0.325
NO-TL:CI42-MLP	5% (30 days)	0.925	0.936	0.254	0.223	10.682	9.102	0.331	0.291
TL: CI42-LSTM-FT	10% (60 days)	0.926	0.944	0.255	0.194	10.775	8.105	0.332	0.260
	20% (120 days)	0.925	**0.948**	0.256	**0.175**	10.946	**7.133**	0.333	**0.238**
M2	1% (6 days)	0.767	0.761	0.492	0.588	18.133	21.604	0.632	0.730
NO-TL:CI42-SVR	5% (30 days)	0.764	0.767	0.495	0.456	18.189	19.388	0.636	0.575
TL: CR12-LSTM-FT	10% (60 days)	0.768	0.798	0.495	0.407	18.225	16.817	0.637	0.524
	20% (120 days)	0.766	**0.813**	0.494	**0.351**	18.267	**14.028**	0.634	**0.467**
M1	1% (6 days)	0.669	0.665	0.612	0.705	21.837	25.636	0.780	0.881
NO-TL:CR12-SVR	5% (30 days)	0.665	0.667	0.614	0.639	21.882	23.484	0.782	0.811
TL: CR12-LSTM-Fr	10% (60 days)	0.673	0.675	0.613	0.544	21.977	19.975	0.782	0.709
	20% (120 days)	0.675	**0.679**	0.612	**0.483**	22.005	**18.410**	0.778	**0.642**

pending on the local weather conditions, the contribution of RH can vary greatly. Consistent with previous research [20,23], RH demonstrates a particularly critical role in humid climatic regions, where moisture interactions significantly influence the evapotranspiration process.

The 12 No-TL models (RF, SVR and MLP) with the lower MAE are selected to be compared with 12 TL models (LSTM) in the next phases.

3.2. ET_0 estimation and forecasting at regional scale

After the local validations, each model is retrained with its entire data set (training and validation) and then used to predict the test set.

Table 4 presents the best mean regional estimation results among all No-TL and TL models. When only 6 days (1% of all test data) are used, all TL models perform worse than No-TL models. However, with more data used to retrain the model, TL models start to outperform No-TL models in all cases. Three out of four TL models use the FT method, which suggests changing the weight of each layer to fit the new data yields better results. A gradual deterioration in all metrics is observed with the reduction in input variables and also the TLPercentage. With complete variables and using 20% of new data to retrain, CI42-LSTM-FT model achieved 0.116 mm d^{-1} in MAE, and 4.761% in MAPE. Conversely, when Rs is used as the only input variable, CR12-LSTM-Fr achieved 0.483 mm d^{-1} in MAE and 18.41% in MAPE, also using 20% of new data to retrain.

The regional estimation result of each model at the station level is shown in Fig. 5.a, using the statistical indices of MAPE in the green box. The models used are those specified in Table 4. It can be seen that the stations close to the sea, namely LO51, CA91 and CA73, had a great improvement compared to other stations closer to the stations where the models were trained (CR12 and CI42), especially in CA73 where MAPE were reduced by more than half (14.44% to 4.9%). The main reason for this improvement could be that the ET_0 values behave differently in these coastal stations with the selected TL models. Therefore, this indicates that the TL technique is a useful method to overcome this discrepancy.

In the case of the indirect method of ET_0 forecasting, the accuracy of the forecasted climatic variables is crucial. If all the variables were forecasted without error, the forecasted ET_0 could have the same per-

Table 5
Best mean regional ET_0 forecast statistical metrics using 12 No-TL and 12 TL models trained from three stations (CI42, CA91, and CR12). The bold values indicate the best values obtained between No-TL and TL models, with the corresponding TLPercentage used to retrain the TL models. The suffix FT in TL models indicates that the model uses the fine-tuning method and Fr indicates that the model uses the freezing all layers method.

Models	TLPercentage	R^2		MAE (mm d^{-1})		MAPE (%)		RMSE (mm d^{-1})	
		NO-TL	TL	NO-TL	TL	NO-TL	TL	NO-TL	TL
M4	10% (9 days)	0.774	0.772	0.541	**0.407**	21.249	**14.781**	0.680	0.559
NO TL:CI42-SVR	20% (18 days)	**0.776**	0.776	0.534	0.415	21.249	15.448	0.667	0.555
TL:CA91-LSTM-Fr	30% (27 days)	0.763	0.763	0.538	0.414	21.585	15.878	0.674	**0.554**
	40% (36 days)	0.722	0.713	0.540	0.420	21.691	16.265	0.676	0.561
M3	10% (9 days)	0.704	0.738	0.638	0.455	24.132	17.249	0.801	0.607
NO TL:CA91-SVR	20% (18 days)	0.702	**0.752**	0.627	**0.427**	24.012	**16.642**	0.784	**0.565**
TL: CI42-LSTM-FT	30% (27 days)	0.683	0.736	0.630	0.451	24.411	18.990	0.788	0.592
	40% (36 days)	0.628	0.694	0.634	0.452	24.560	18.222	0.789	0.593
M2	10% (9 days)	**0.610**	0.600	0.635	0.573	23.003	22.008	0.814	0.761
NO TL:CI42-SVR	20% (18 days)	0.602	0.606	0.633	**0.535**	22.911	**20.159**	0.808	**0.710**
TL: CI42-LSTM-Fr	30% (27 days)	0.574	0.575	0.640	0.573	23.318	24.019	0.818	0.749
	40% (36 days)	0.522	0.532	0.643	0.592	23.181	24.793	0.814	0.761
M1	10% (9 days)	**0.531**	0.527	0.674	0.635	24.009	**22.949**	0.860	0.831
NO TL:CR12-SVR	20% (18 days)	0.517	0.512	0.674	0.635	24.125	25.474	0.862	0.828
TL: CR12-LSTM-Fr	30% (27 days)	0.496	0.498	0.671	0.643	24.344	29.756	0.860	0.817
	40% (36 days)	0.456	0.458	0.650	**0.623**	23.719	29.681	0.829	**0.790**

Fig. 6. Statistical analysis (MAE and CVRMSE) of WeatherBit in all Murcia stations.

formances archived in the estimation, and the inaccuracy of the input variables (or some of them) could lead to highly biased results even if the model is accurate.

The two statistical errors (MAE and CVRMSE) of WeatherBit are given in Fig. 6. In terms of CVRMSE, the most biased forecast variable is U_2 in all Murcia stations, ranging from 22.83% to 127.84%, with a mean value of 62%, followed by RHmin, Tmin, Rs, RHmax and Tmax, with mean CVRMSE 19.97%, 15.31%, 14.40%, 11.58% and 4.82%, respectively. Although Rs has the maximum errors in MAE (from 17.80 to 26.74 W m^{-2} and a mean of 20.42 W m^{-2}), this is because its range is much wider than the other variables. The mean of the other variables are RHmax (7.15%), RHmin (5.34%), Tmin (1.47 ºC), Tmax (0.93 ºC), and U_2 (0.64 m s^{-1}).

Table 5 shows the best average regional forecast results, and the result of each model at the station level can be observed in Fig. 5.b.

Remarkably, the best performing models are no longer the ones with the highest TLPercentage. Furthermore, in most cases, with more new data to retrain the models, their performance is worse, representing a possible overfitting problem where the models are trying to approximate the real values with forecasted climatic variables, where most of them contain different levels of errors. Therefore, with more retraining, the models may have more errors.

In terms of MAE, for the M4, M3, and M2 models, the best option for retraining is to use 10%, 10%, and 20%, respectively. It is only for the M1 that using 40% gives a better performance. The R^2 values are significantly reduced in all models, meaning that the models are less effective at capturing the total variability in the data. Regarding the rest

Table 6
Best mean national ET_0 estimation statistical metrics using 12 No-TL and 12 TL models trained from three stations (CI42, CA91, and CR12). The bold values indicate the best values obtained between No-TL and TL models, with the corresponding TLPercentage used to retrain the TL models. The suffix FT in TL models indicates that the model uses the fine-tuning method and Fr indicates that the model uses the freezing all layers method.

Models	TLPercentage	R^2		MAE (mm d^{-1})		MAPE (%)		RMSE (mm d^{-1})	
		NO-TL	TL	NO-TL	TL	NO-TL	TL	NO-TL	TL
M4	1% (6 days)	0.957	0.963	0.204	0.232	10.922	12.812	0.261	0.281
NO TL:CR12-SVR	5% (30 days)	0.956	0.964	0.205	0.197	11.015	10.632	0.261	0.245
TL: CR12-LSTM-FT	10% (60 days)	0.956	0.969	0.205	0.157	11.169	8.372	0.262	0.202
	20% (120 days)	0.955	**0.976**	0.206	**0.117**	11.197	**6.280**	0.262	**0.153**
M3	1% (6 days)	0.912	0.909	0.355	0.299	18.820	18.363	0.431	0.376
NO TL:CR12-SVR	5% (30 days)	0.910	0.916	0.355	0.248	18.956	13.587	0.432	0.322
TL: CI42-LSTM-FT	10% (60 days)	0.911	0.930	0.355	0.218	19.198	12.144	0.432	0.286
	20% (120 days)	0.908	**0.940**	0.357	**0.179**	19.235	**9.373**	0.435	**0.241**
M2	1% (6 days)	0.787	0.805	0.445	0.469	19.635	23.631	0.574	0.583
NO TL:CI42-SVR	5% (30 days)	0.785	0.818	0.442	0.394	19.670	19.723	0.570	0.501
TL: CR12-LSTM-FT	10% (60 days)	0.790	0.839	0.440	0.353	19.852	17.328	0.568	0.455
	20% (120 days)	0.784	**0.850**	0.440	**0.314**	19.703	**15.090**	0.567	**0.409**
M1	1% (6 days)	0.688	0.685	0.581	0.698	27.109	33.643	0.740	0.863
NO TL:CR12-SVR	5% (30 days)	0.687	0.685	0.577	0.579	27.132	28.979	0.735	0.736
TL: CR12-LSTM-Fr	10% (60 days)	**0.693**	0.691	0.575	0.521	27.394	23.910	0.733	0.679
	20% (120 days)	0.688	0.690	0.574	**0.454**	27.139	**20.657**	0.732	**0.602**

Table 7
Best mean national ET_0 forecasting statistical metrics using 12 No-TL and 12 TL models trained from three stations (CI42, CA91, and CR12). The bold values indicate the best values obtained between No-TL and TL models, with the corresponding TLPercentage used to retrain the TL models. The suffix FT in TL models indicates that the model uses the fine-tuning method and Fr indicates that the model uses the freezing all layers method.

Models	TLPercentage	R^2		MAE (mm d^{-1})		MAPE (%)		RMSE (mm d^{-1})	
		NO-TL	TL	NO-TL	TL	NO-TL	TL	NO-TL	TL
M4	10% (9 days)	0.718	**0.755**	0.855	0.485	41.753	23.035	0.995	0.615
NO TL: CI42-SVR	20% (18 days)	0.698	0.740	0.839	0.475	41.709	22.903	0.972	0.599
TL: CA91-LSTM-Fr	30% (27 days)	0.686	0.727	0.826	0.461	41.440	**22.602**	0.959	**0.581**
	40% (36 days)	0.652	0.692	0.826	**0.460**	42.144	23.058	0.956	0.581
M3	10% (9 days)	0.472	**0.702**	1.030	0.530	52.313	26.233	1.226	0.674
NO TL: CA91-SVR	20% (18 days)	0.461	0.685	1.016	0.492	52.676	26.049	1.205	0.629
TL: CR12-LSTM-FT	30% (27 days)	0.450	0.684	1.008	0.440	52.837	21.575	1.197	0.566
	40% (36 days)	0.434	0.664	1.018	**0.434**	54.446	**21.055**	1.202	**0.558**
M2	10% (9 days)	0.638	**0.649**	0.581	0.580	**25.692**	31.430	0.717	0.722
NO TL: CA91-SVR	20% (18 days)	0.582	0.600	0.570	0.545	25.730	30.089	0.702	0.682
TL: CA91-LSTM-Fr	30% (27 days)	0.561	0.578	0.570	0.531	25.955	30.396	0.703	0.665
	40% (36 days)	0.549	0.572	0.571	**0.507**	26.576	27.531	0.700	**0.635**
M1	10% (9 days)	**0.533**	0.527	0.655	0.674	**28.046**	32.159	0.821	0.836
NO TL: CA91-SVR	20% (18 days)	0.479	0.474	0.643	0.654	28.085	34.481	0.803	0.812
TL: CR12-LSTM-Fr	30% (27 days)	0.464	0.464	0.638	0.637	28.341	36.966	0.799	0.787
	40% (36 days)	0.452	0.456	**0.626**	0.633	28.824	39.096	0.783	**0.777**

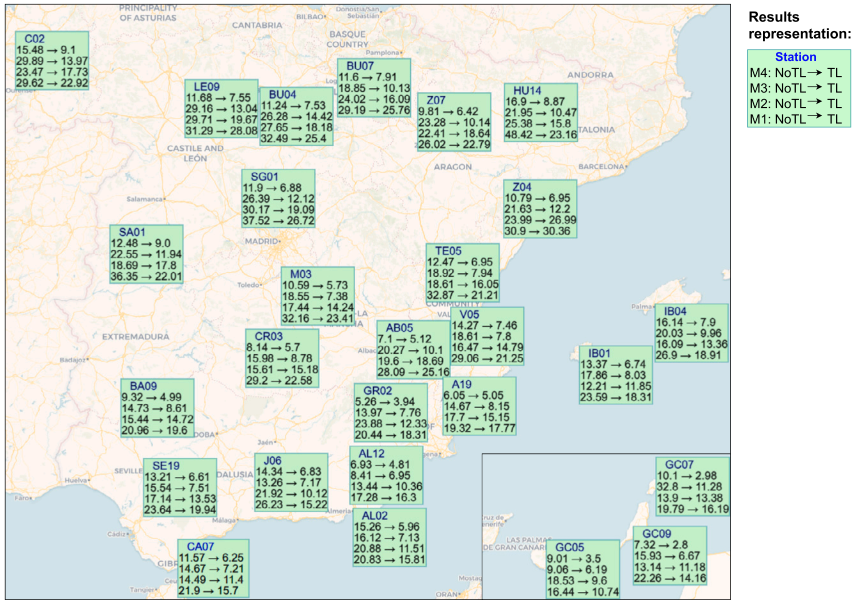
of the metrics, the best four models have values ranging from 14.781% to 22.949% in MAPE and from 0.554 mm d^{-1} to 0.790 mm d^{-1} in RMSE. In contrast, three out of four TL models now use the Fr method, which means that using TL process as a feature extraction method is more suitable in this case, also trying to avoid the overfitting problem, as we mentioned before.

Finally, it should be noted that despite the smaller improvement in forecasting compared to estimation by TL models, they still achieve a significant enhancement compared to No-TL models (reduction of the MAE ranged from 0.027 to 0.199 mm d^{-1}). In particular, at the station level, with the exception of some M1 and M2 models that use only Rs or Rs and T as inputs and therefore their performance can vary widely (i.e. be more unstable), most TL models showed great improvements. Again, CA91 and CA73 had the most significant reductions.

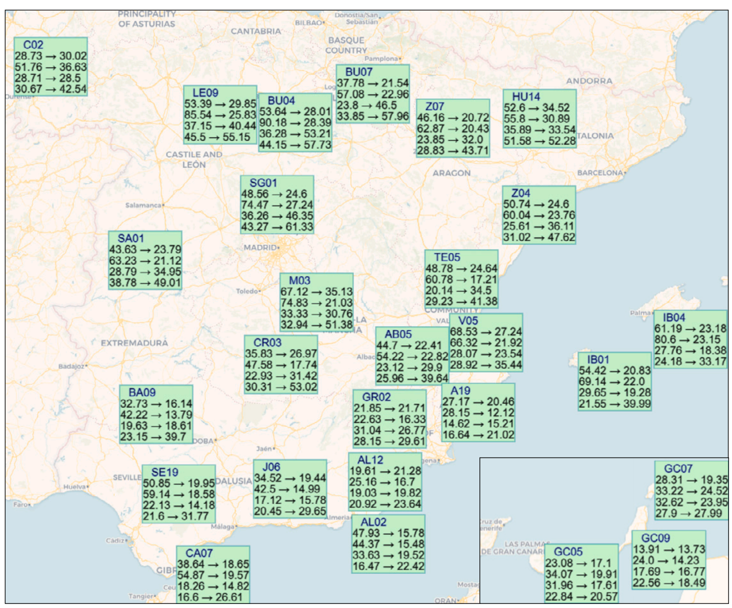
3.3. ET_0 estimation and forecasting at national scale

Estimating and forecasting ET_0 at the national level follows the same process as at the regional level. Table 6 summarizes the four average statistical indicators of the No-TL and TL models with the same structure as before at the national ET_0 estimation. The results once again demonstrate the effectiveness of the TL method, where all TL models achieve similar performances to those obtained at the regional level with the same configurations, i.e. the more data used to retrain the model the better, and three out of four of the TL models have used the fine-tuning method. Specifically, using 20% of the data to retrain the models, the CR12-LSTM-FT (M4) has the best performance (MAE = 0.117 mm d^{-1} and MAPE = 6.28%), the CI42-LSTM-FT (M3) has a slight decrease (MAE = 0.179 mm d^{-1} and MAPE = 9.373%), then the CR12-LSTM-FT (M2) shows a noticeable reduction (MAE = 0.314 mm d^{-1} and MAPE = 15.09%). Finally, the CR12-LSTM-Fr (M1) shows the lowest overall R^2 values, with a maximum value of 0.693 (10% of TLPercentage). However, the improvement is still significant with 20% of TLPercentage (MAE = 0.454 mm d^{-1} and MAPE = 20.657%).

The improvement of each station by the TL models is shown in Fig. 7.a. The stations with the major improvement are those in the Canary Islands, where the climate is very different and the variation of ET_0 between seasons is much smaller. The M2 model at station GC07



(a) Spain ET_0 estimation improvements using TL



(b) Spain ET_0 forecast improvements using TL

Fig. 7. Comparison of national ET_0 estimation and forecast results between the best performing No-TL and TL models. The green box represents the results (MAPE) in the corresponding station.

had the greatest improvement, with a reduction of 21.52% in MAPE. In the remaining stations, the improvement varies widely, possibly due to the microclimate of the individual zones, but overall, all stations show a general improvement.

When it comes to forecasting, Fig. 8 describes the MAE and CVRMSE of NOAA GFS with the real data. The U_2 is still the most biased, followed by Tmin, RHmin, Rs, RHmax and Tmax, the mean CVRMSE and MAE are 111.15% (1.25 m s^{-1}), 39.61% (2.88 ºC), 25.45% (7.81%), 17.64% (23.61 W m^{-1}), 17.59% (12.53%) and 9% (1.68 ºC), respectively.

Table 7 shows the national ET_0 forecast results. Even with a greater number of stations and varying climatic conditions, the TL models maintain similar performance at the national scale as in the regional forecast, as measured by R^2, MAE, and RMSE. Remarkably, the M3 No-TL models had the largest error before the TL process, the main reason is because two important input variables, i.e. Tmin and U_2, were two worst predicted variables by the NOAA GFS, therefore the M3 No-TL models used very biased input information to make the forecast. However, after the retraining process, the TL model has learned and adapted to this error. Hence, after the TL process, the CR12-LSTM-FT (M3) had the smallest errors in MAE (0.434 mm d^{-1}), MAPE (21.055%), and RMSE (0.558 mm d^{-1}). The significant performance improvement stems primarily from resolving the U_2 bias in NWP services, not from U_2's intrinsic variable importance. Regarding the M2 and M1 models, they became more unstable. In terms of MAPE, the TL gave worse results, although the RMSE showed a consistent improvement for all models. Once again, three of the best four TL models used the freeze method, as did the regional forecast.

At the station level (Fig. 7.b), M3 model with TL process improved performance in all stations, similar to the M4 model with two exceptions (C02 and AL12). Nevertheless, most of the M2 and M1 models had only a slight reduction or even worse performance in most cases. It is important to highlight that even with all input variables, most of the M4 models

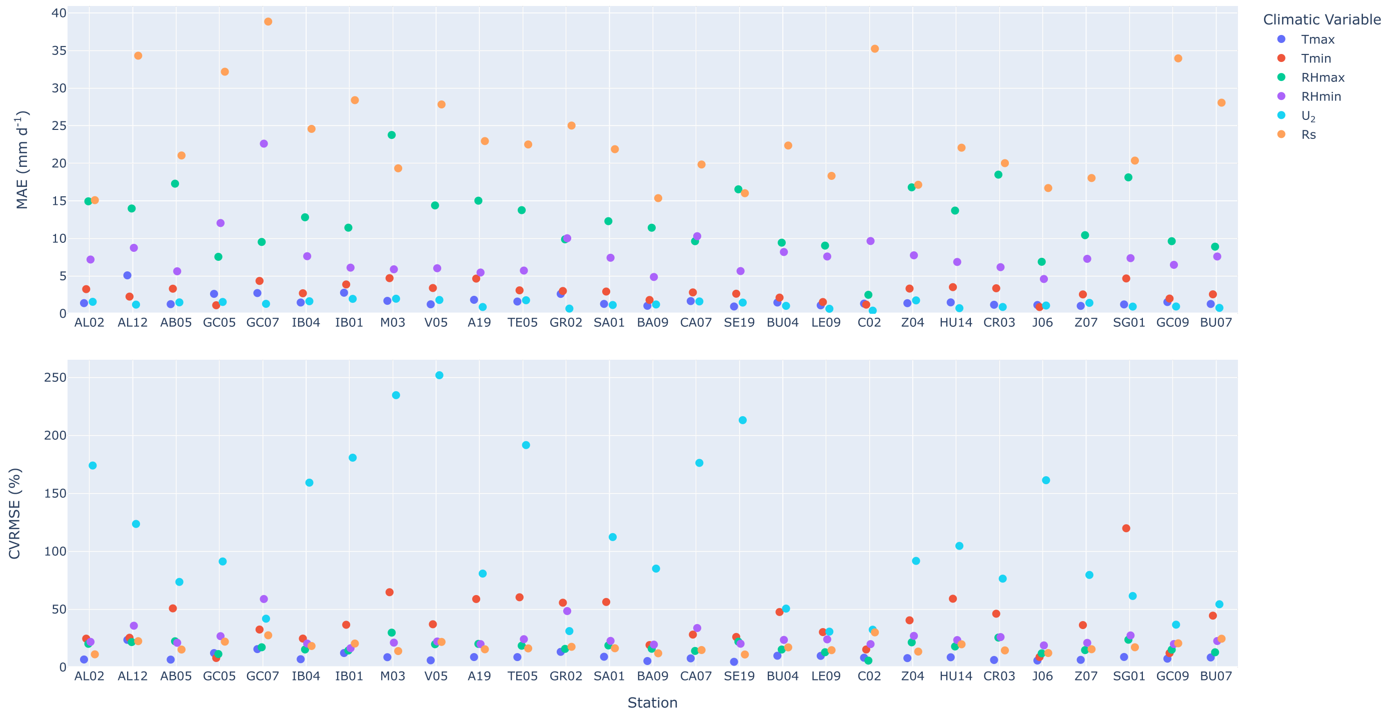


Fig. 8. Statistical analysis (MAE and CVRMSE) of NOAA GFS in all Spain stations.

without TL process had about 41% of errors, which is inappropriate to be used in real scenarios. After the TL process, the average MAPE becomes 22% and it is acceptable as a forecasting method for the whole nation. This effect is even more pronounced for the M3 model (54% to 21%).

4. Discussion

4.1. Summary and comparison with previous studies

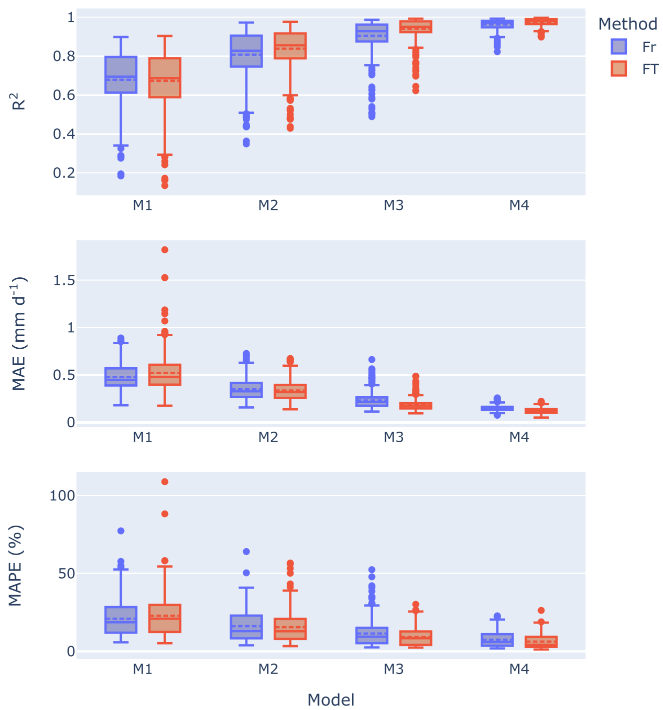
In the first place, at the local scale, all input combinations were evaluated. In terms of MAE, which is the criterion chosen to rank the estimation performance of the models, the best combination for each model (M1 to M4) is the same, and SVR outperformed the other algorithms.

In general, the models archived at the local stations gave excellent estimation results. In the work with similar objectives, Zhu et al. [52] developed hybrid Particle Swarm Optimization (PSO) with ELM, ELM, ANN, and RF at four stations in the arid region of northwest China. The PSO-ELM model provided the best results, with R^2 ranging from 0.886-0.969 and MAE from 0.268-0.536 mm d^{-1} using radiation-based inputs, which are: Tmax, Tmin, extraterrestrial radiation (Ra), and Rs. These inputs correspond to our M2 models (without using Ra), with R^2 ranging from 0.953-0.976 and MAE ranging from 0.208-0.337 mm d^{-1}, which are slightly improved. Yamaç [48] used a 20-year data set from the station in the Middle Anatolia region of Turkey, applying KNN and ANN to estimate daily ET_0 with the same M4 input in this study. The ANN model gave the result with $R^2 = 0.991$, MAE = 0.162 mm d^{-1} and RMSE = 0.225 mm d^{-1}, which are very close to our results with CA91 M4, also implemented with ANN (MLP with $R^2 = 0.987$, MAE = 0.138 mm d^{-1} and RMSE = 0.208 mm d^{-1}).

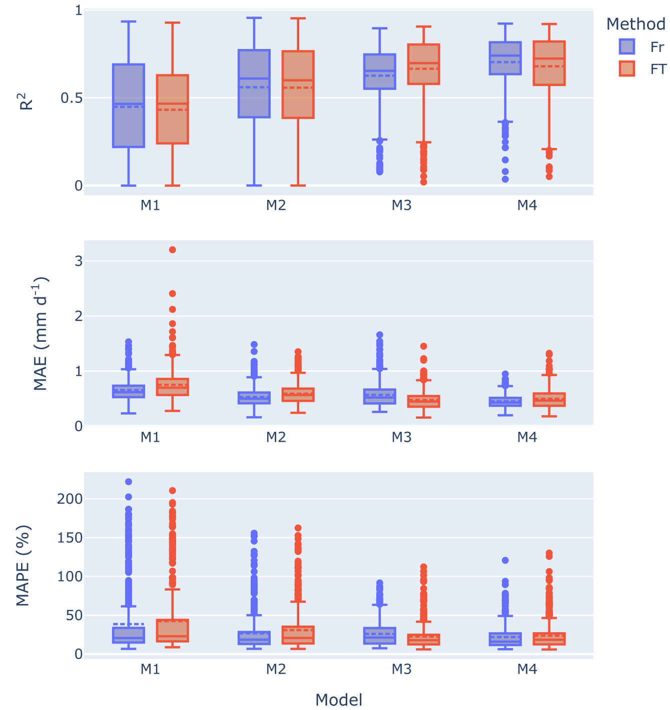
At regional and national scale, seeing the overall estimation results, both M4 and M3 TL models are suitable to replace the FAO56-PM equation to estimate the ET_0 of the region of Murcia. If only T and Rs are available, the M2 and M1 models can also be used to obtain the ET_0 trend or to impute the missing values. With 20% of new data to retrain, the percentage improvement in MAE is 54.3% at the regional level and 99.5% at the national level.

In the work of Zhu et al. [52], they have also tested the models over five stations, being the PSO-ELM again the best-performing model with average $R^2 = 0.935$ and MAE = 0.470 mm d^{-1}, using the same radiation-based input. In Spain, Bellido-Jiménez et al. [9] implemented a regional ML method that collected datasets from 89 AWS in southern regions to estimate their ET_0 using MLP and ELM. The MLP model had the best results with the mean $R^2 = 0.922$ and RMSE = 0.669 mm d^{-1} using the Ra, Tmax from the previous and same day, Tmin from the same day, the time when Tmax occurs from the same day and the integral of the half-hourly temperature records from the previous and same day as inputs. Both studies used similar inputs to our M2 models, which gave better performance in terms of MAE (0.351 mm d^{-1} and 0.314 mm d^{-1} at the regional and national level, respectively) and RMSE (0.467 mm d^{-1} and 0.409 mm d^{-1} at the regional and national level, respectively). However, our model performed worse with R^2 (0.813 and 0.85 at the regional and national level, respectively). This may seem contradictory at first, but it is important to understand how R^2 is interpreted. R^2 measures the proportion of the total variance in the data that the model can explain, and it is calculated by comparing the sum of squares of the residuals (errors) to the total sum of squares (the total variability of the data). A lower R^2 does not necessarily mean that the model is worse in terms of absolute accuracy; it could be penalized for a slightly poorer fit in terms of capturing the total variability in the data. In addition, R^2 is sensitive to how errors are distributed over different parts of the data range. It is possible that our model fits extreme values (very low or very high) better, but does not capture the total variability of the data as well as the model of other studies. Therefore, although the absolute errors are smaller (as reflected in the MAE and RMSE), the model may not capture the general trend or structure of the data as well, which is penalized in the R^2 calculation.

On the other hand, the main problem identified in the forecast, both regionally and nationally, was the highly biased forecast of climatic variables by the NWP services, mainly U_2, RHmin, and Tmin. Nevertheless, the TL process had a very positive impact in correcting these input errors. At the regional scale, the percentage improvement in MAE is ranged from 4.29% to 46.62%, and at the national scale from -2.75% to 134.27%.



(a) Estimation



(b) Forecasting

Fig. 9. Boxplots of the comparison between FT and Fr methods (R^2, MAE and MAPE) during the TL process across all training stations.

Comparing with works with similar methodology, Fan et al. [13] conducted a study to forecast ET_0 up to 16 days lead time across 51 weather stations in various climatic zones of China using the four forecast climatic variables, which were collected from the NOAA Global Ensemble Reforecast v2 Data. They also concluded that U_2 had the worst forecast performance and the average R^2 values of ET_0 forecast varied from 0.56 to 0.85 and the average RMSE from 0.6 to 1.0 mm d^{-1}. In the work of Perera et al. [36] in Austria, the mean RMSE and R^2 of forecasted ET_0 were 0.73 mm d^{-1} and 0.85 for 1 day lead time, respectively. In this study, the TL models (M4) archived similar performance in R^2 values (0.776 at regional scale and 0.755 at national scale) and better RMSE (0.554 mm d^{-1} and 0.581 mm d^{-1} at regional and national scale, respectively).

Finally, regarding the use of the FT or Fr method during the TL process, Fig. 9 describes the comparison of two methods in estimation and in forecasting, including all the results of the three training stations. It can be seen that the results are consistent with those shown in Table 4, Table 5, Table 6 and Table 7, where FT has more compact and less errors in estimation (except for M1 models) and Fr has better performance in forecasting (except for M3 models). Our interpretation of this fact is that when the inputs are absent of errors (case of estimation with real data), adjusting the weight of each neuron to the new information can yield better results. In contrast, when the inputs contain highly biased data (case of forecasting with NWP data), maintaining the original weights with learned features is more reliable.

4.2. Seasonal analysis

The performance of the TL models in different seasons over all stations (regional and national) is presented in Fig. 10.

In the estimation, in terms of R^2 and MAPE, autumn had the worst results, with the highest mean and maximum values in MAPE for all TL models, followed by winter, while spring had the best performance, then summer. However, in terms of MAE, all M1 and M2 TL models had contrasting results, with spring and summer having larger errors than autumn and winter. This is due to the fact that the range of ET_0 is much wider in spring and summer than in the other two seasons, then the lack of input information causes larger errors.

In the forecast, autumn was still the worst season, especially the M1 model, which had an average $R^2 = 0.117$, MAE $= 0.731$ mm d^{-1}, and MAPE $= 91.24\%$. Winter was also worse than spring and summer in terms of R^2 and MAPE. However, the difference in MAE for models M2 to M4 across spring, summer, and winter was not significant (mean ranged from 0.439 mm d^{-1} to 0.653 mm d^{-1}).

5. Conclusion

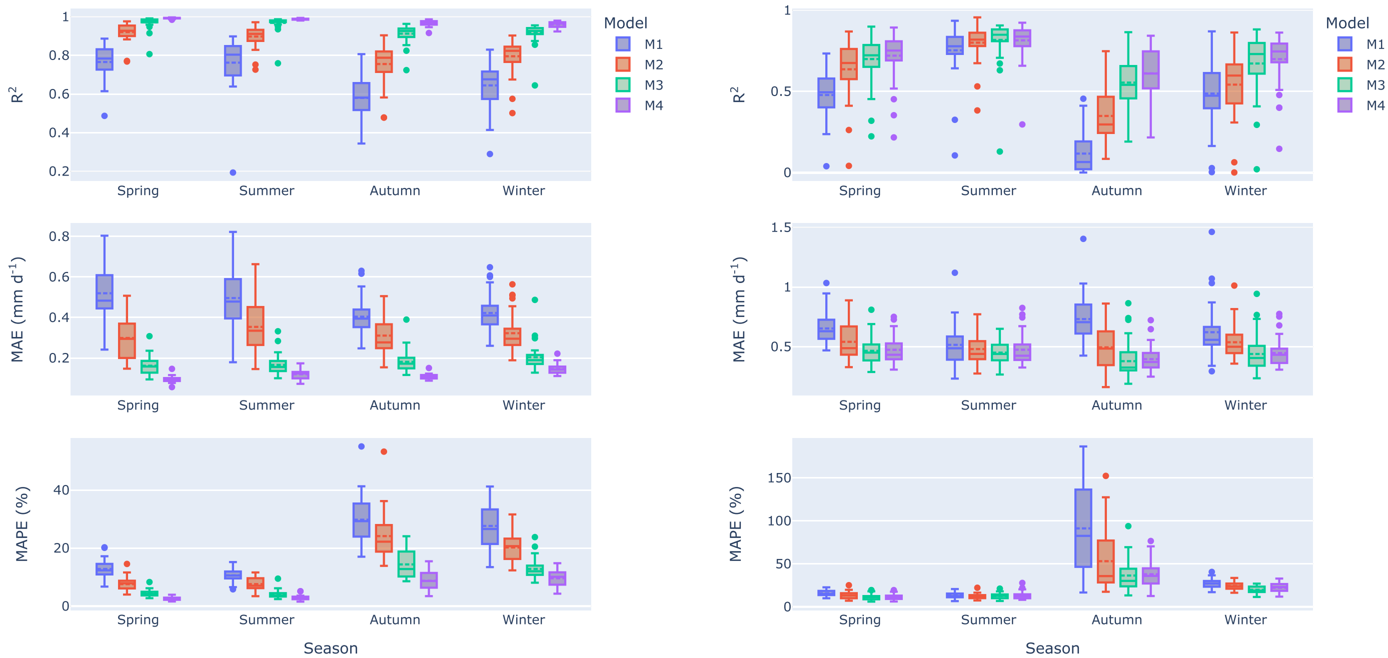
In this study, the main objective is to have a feasible alternative methodology of FAO56-PM through ML algorithms with the requirement of obtaining as much generalizability as possible. In this sense, as a novel approach in the field of ET_0 computation, TL has been used to retrain the DL models in order to compare the models with and without TL process over different climatic conditions and spatial scales of Spain.

At the local level, according to the validation results, Rs has the main contribution to the ET_0 computation, followed by T, U_2 and RH. As for the ML algorithms, SVR outperformed LSTM, MLP, and RF in most cases.

Regarding the estimation, both regional and national scale showed that with more than 5% of new data (30 days), there is a continuous improvement for TL models. The TL process is recommended for all models, and the M4/M3 models are recommended for use in real scenarios since their average MAPE can be less than 10%. In the case of limited variables, the M2/M1 models can be used as a tool to obtain the ET_0 trend or to impute the missing values.

In forecasting, U_2 is identified as the most biased climatic variable obtained from NWP, and contrary to the estimation, TL models showed improvements from the beginning. In this case, TL process is only recommended to be applied for M4/M3 models (average MAPE ranged from 14.78% to 22.6%) and not for M2/M1 due to their unstable performance.

Finally, with the increasing development of digital twin technologies, as one of the most important parameters for agricultural activities, ET_0 is essential for any related digital twin in agriculture. Therefore,



(a) Estimation (b) Forecasting

Fig. 10. Boxplots of the seasonal statistical analysis (R^2, MAE and MAPE) of the best performing TL models across all stations.

the integration of the TL technique and the final models obtained in this study into a developing digital twin [50] to improve its decision-making capability is considered as an ongoing work.

CRediT authorship contribution statement

Yu Ye: Writing – review & editing, Writing – original draft, Software, Investigation, Conceptualization. **Aurora González-Vidal:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation. **Miguel A. Zamora-Izquierdo:** Writing – review & editing, Methodology, Conceptualization. **Antonio F. Skarmeta:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

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.

Acknowledgement

This study was supported by grant RYC2023-043553-I, funded by MICIU/AEI/10.13039/501100011033 and European Social Fund. It was also supported by the projects of the Innovation Spanish Agency: EDEN (CPP2021-009146) and DEMETER 5.0 (PLEC2022-009289). It was partially funded by the HORIZON-MSCA-2021-SE-01-01 project Cloudstars (g.a. 101086248) as well as by projects GEMINI TED2021-129767B-I00 and financed by MCIN/AEI/10.130 39/501100011033 and the European Union NextGenerationEU/PRTR.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.atech.2025.100886.

Data availability

I have shared the link to my data/code in the paper.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, et al., TensorFlow: large-scale machine learning on heterogeneous systems, software available from tensorflow.org, https://www.tensorflow.org/, 2015.

[2] P. Aghelpour, R. Norooz-Valashedi, Predicting daily reference evapotranspiration rates in a humid region, comparison of seven various data-based predictor models, Stoch. Environ. Res. Risk Assess. 36 (2022) 4133–4155.

[3] M.M. Alam, M.Y. Akter, A.R.M.T. Islam, J. Mallick, Z. Kabir, R. Chu, A. Arabameri, S.C. Pal, M.A. Al Masud, R. Costache, et al., A review of recent advances and future prospects in calculation of reference evapotranspiration in Bangladesh using soft computing models, J. Environ. Manag. 351 (2024) 119714.

[4] R.G. Allen, W.O. Pruitt, J.L. Wright, T.A. Howell, F. Ventura, R. Snyder, D. Itenfisu, P. Steduto, J. Berengena, J.B. Yrisarry, M. Smith, L.S. Pereira, D. Raes, A. Perrier, I. Alves, I. Walter, R. Elliott, A recommendation on standardized surface resistance for hourl2y calculation of reference eto by the fao56 Penman-Monteith method, Agric. Water Manag. 81 (2006) 1–22, https://doi.org/10.1016/j.agwat.2005.03.007.

[5] R.G. Allen, D. Raes, M. Smith, Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56, 1998.

[6] J. Almorox, V.H. Quej, P. Martí, Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes, J. Hydrol. 528 (2015) 514–522, https://doi.org/10.1016/j.jhydrol.2015.06.057.

[7] H.E. Beck, N.E. Zimmermann, T.R. McVicar, N. Vergopolan, A. Berg, E.F. Wood, Present and future Köppen-Geiger climate classification maps at 1-km resolution, Sci. Data 5 (2018) 1–12.

[8] J.A. Bellido-Jiménez, J. Estévez, A.P. García-Marín, New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain, Agric. Water Manag. 245 (2021) 106558, https://doi.org/10.1016/j.agwat.2020.106558.

[9] J.A. Bellido-Jiménez, J. Estévez, A.P. García-Marín, A regional machine learning method to outperform temperature-based reference evapotranspiration estimations in southern Spain, Agric. Water Manag. 274 (2022) 107955, https://doi.org/10.1016/j.agwat.2022.107955.

[10] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[11] F. Chollet, et al., Keras, https://keras.io, 2015.

[12] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.

[13] J. Fan, L. Wu, J. Zheng, F. Zhang, Medium-range forecasting of daily reference evapotranspiration across China using numerical weather prediction outputs downscaled by extreme gradient boosting, J. Hydrol. 601 (2021) 126664, https://doi.org/10.1016/j.jhydrol.2021.126664.

[14] FAO, Water for sustainable food and agriculture: a report produced for the G20 presidency of Germany, https://www.fao.org/3/i7959e/i7959e.pdf, 2017.

[15] Y. Feng, N. Cui, D. Gong, Q. Zhang, L. Zhao, Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling, Agric. Water Manag. 193 (2017) 163–173.

[16] L.B. Ferreira, F.F. da Cunha, Multi-step ahead forecasting of daily reference evapotranspiration using deep learning, Comput. Electron. Agric. 178 (2020) 105728, https://doi.org/10.1016/j.compag.2020.105728.

[17] L.B. Ferreira, F.F. da Cunha, R.A. de Oliveira, E.I. Fernandes Filho, Estimation of reference evapotranspiration in Brazil with limited meteorological data using ann and svm – a new approach, J. Hydrol. 572 (2019) 556–570, https://doi.org/10.1016/j.jhydrol.2019.03.028.

[18] M. Gocić, S. Motamedi, S. Shamshirband, D. Petković, S. Ch, R. Hashim, M. Arif, Soft computing approaches for forecasting reference evapotranspiration, Comput. Electron. Agric. 113 (2015) 164–173, https://doi.org/10.1016/j.compag.2015.02.010.

[19] A. González-Vidal, P. Rathore, A.S. Rao, J. Mendoza-Bernal, M. Palaniswami, A.F. Skarmeta-Gómez, Missing data imputation with Bayesian maximum entropy for Internet of things applications, IEEE Internet Things J. 8 (2020) 16108–16120.

[20] P. Goyal, S. Kumar, R. Sharda, A review of the artificial intelligence (AI) based techniques for estimating reference evapotranspiration: current trends and future perspectives, Comput. Electron. Agric. 209 (2023) 107836, https://doi.org/10.1016/j.compag.2023.107836, https://www.sciencedirect.com/science/article/pii/S0168169923002247.

[21] F. Granata, F. Di Nunno, Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks, Agric. Water Manag. 255 (2021) 107040, https://doi.org/10.1016/j.agwat.2021.107040.

[22] Z. Hu, R.N. Bashir, A.U. Rehman, S.I. Iqbal, M.M.A. Shahid, T. Xu, Machine learning based prediction of reference evapotranspiration (et0) using IoT, IEEE Access 10 (2022) 70526–70540, https://doi.org/10.1109/ACCESS.2022.3187528.

[23] G. Huang, L. Wu, X. Ma, W. Zhang, J. Fan, X. Yu, W. Zeng, H. Zhou, Evaluation of catboost method for prediction of reference evapotranspiration in humid regions, J. Hydrol. 574 (2019), https://doi.org/10.1016/j.jhydrol.2019.04.085.

[24] A. Iglesias, C. Rosenzweig, D. Pereira, Agricultural impacts of climate change in Spain: developing tools for a spatial analysis, Glob. Environ. Change 10 (2000) 69–80, https://doi.org/10.1016/S0959-3780(00)00010-8, https://www.sciencedirect.com/science/article/pii/S0959378000000108.

[25] M. Iman, H.R. Arabnia, K. Rasheed, A review of deep transfer learning and recent advancements, Technologies 11 (2023) 40.

[26] Y.Z. Kaya, M. Zelenakova, F. Üneş, M. Demirci, H. Hlavata, P. Mesaros, Estimation of daily evapotranspiration in Košice city (Slovakia) using several soft computing techniques, Theor. Appl. Climatol. 144 (2021) 287–298.

[27] D. Kool, N. Agam, N. Lazarovitch, J. Heitman, T. Sauer, A. Ben-Gal, A review of approaches for evapotranspiration partitioning, Agric. For. Meteorol. 184 (2014) 56–70, https://doi.org/10.1016/j.agrformet.2013.09.003.

[28] B. Liu, M. Liu, Y. Cui, D. Shao, Z. Mao, L. Zhang, S. Khan, Y. Luo, Assessing forecasting performance of daily reference evapotranspiration using public weather forecast and numerical weather prediction, J. Hydrol. 590 (2020) 125547, https://doi.org/10.1016/j.jhydrol.2020.125547.

[29] X. Liu, L. Wu, F. Zhang, G. Huang, F. Yan, W. Bai, Splitting and length of years for improving tree-based models to predict reference crop evapotranspiration in the humid regions of China, Water 13 (2021) 3478.

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

[31] H. Medina, D. Tian, P. Srivastava, A. Pelosi, G.B. Chirico, Medium-range reference evapotranspiration forecasts for the contiguous United States based on multi-model numerical weather predictions, J. Hydrol. 562 (2018) 502–517, https://doi.org/10.1016/j.jhydrol.2018.05.029.

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

[33] P. Open-Meteo Zippenfenig, Open-meteo.com weather api, https://open-meteo.com/, 2023, https://doi.org/10.5281/zenodo.7970649.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[35] L.S. Pereira, R.G. Allen, M. Smith, D. Raes, Crop evapotranspiration estimation with fao56: past and future, Agric. Water Manag. 147 (2015) 4–20, https://doi.org/10.1016/j.agwat.2014.07.031, Agricultural Water Management: Priorities and Challenges.

[36] K.C. Perera, A.W. Western, B. Nawarathna, B. George, Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs, Agric. For. Meteorol. 194 (2014) 50–63, https://doi.org/10.1016/j.agrformet.2014.03.014.

[37] S.M. Ravindran, S.K.M. Bhaskaran, S.K.N. Ambat, A deep neural network architecture to model reference evapotranspiration using a single input meteorological parameter, Environ. Process. 8 (2021) 1567–1599.

[38] S.M. Ribeiro, C.L. de Castro, Missing data in time series: a review of imputation methods and case study, in: Learning and Nonlinear Models-Revista da Sociedade Brasileira de Redes Neurais-Special Issue, in: Time Series Analysis and Forecasting Using Computational Intelligence, vol. 19, 2021.

[39] J. Schmidhuber, S. Hochreiter, et al., Long short-term memory, Neural Comput. 9 (1997) 1735–1780.

[40] SciKeras, A. Garcia Badaracco, Scikeras, https://github.com/adriangb/scikeras, 2024.

[41] N. Shrestha, S. Shukla, Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment, Agric. For. Meteorol. 200 (2015) 172–184, https://doi.org/10.1016/j.agrformet.2014.09.025.

[42] A.A. Suleiman, G. Hoogenboom, Comparison of Priestley-Taylor and fao-56 Penman-Monteith for daily reference evapotranspiration estimation in Georgia, J. Irrig. Drain. Eng. 133 (2007) 175–182.

[43] Y. Tikhamarine, A. Malik, D. Souag-Gamane, O. Kisi, Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration, Environ. Sci. Pollut. Res. 27 (2020) 30001–30019.

[44] A.F. Torres, W.R. Walker, M. McKee, Forecasting daily potential evapotranspiration using machine learning and limited climatic data, Agric. Water Manag. 98 (2011) 553–562, https://doi.org/10.1016/j.agwat.2010.10.012.

[45] S. Traore, Y. Luo, G. Fipps, Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages, Agric. Water Manag. 163 (2016) 363–379, https://doi.org/10.1016/j.agwat.2015.10.009.

[46] J.D. Valiantzas, Simplified forms for the standardized fao-56 penman–monteith reference evapotranspiration using limited weather data, J. Hydrol. 505 (2013) 13–23, https://doi.org/10.1016/j.jhydrol.2013.09.005.

[47] WeatherBit, Weatherbit.io, https://www.weatherbit.io/, 2023.

[48] S.S. Yamaç, Reference evapotranspiration estimation with knn and ann models using different climate input combinations in the semi-arid environment, J. Agric. Sci. (2021).

[49] Y. Yang, Y. Cui, K. Bai, T. Luo, J. Dai, W. Wang, Y. Luo, Short-term forecasting of daily reference evapotranspiration using the reduced-set Penman-Monteith model and public weather forecasts, Agric. Water Manag. 211 (2019) 70–80, https://doi.org/10.1016/j.agwat.2018.09.036.

[50] Y. Ye, A. González-Vidal, A. Cisterna-García, A. Pérez-Ruzafa, M.A.Z. Izquierdo, A.F. Skarmeta, Advancing towards a marine digital twin platform: modeling the Mar Menor coastal lagoon ecosystem in the south western Mediterranean, arXiv preprint, arXiv:2409.10134, 2024.

[51] Y. Yin, S. Wu, D. Zheng, Q. Yang, Radiation calibration of fao56 penman–monteith model to estimate reference crop evapotranspiration in China, Agric. Water Manag. 95 (2008) 77–84, https://doi.org/10.1016/j.agwat.2007.09.002.

[52] B. Zhu, Y. Feng, D. Gong, S. Jiang, L. Zhao, N. Cui, Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data, Comput. Electron. Agric. 173 (2020) 105430, https://doi.org/10.1016/j.compag.2020.105430, https://www.sciencedirect.com/science/article/pii/S0168169919320472.