

**LUCAS BORGES FERREIRA**

**APPLICATIONS OF MACHINE LEARNING FOR REFERENCE  
EVAPOTRANSPIRATION MODELING**

Thesis submitted to the Agricultural Engineering  
Graduate Program of the Universidade Federal de  
Viçosa in partial fulfillment of the requirements  
for the degree of *Doctor Scientiae*.

Advisor: Fernando França da Cunha

**VIÇOSA - MINAS GERAIS  
2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

F383a  
2022  
Ferreira, Lucas Borges, 1993-  
Applications of machine learning for reference  
evapotranspiration modeling / Lucas Borges Ferreira. – Viçosa,  
MG, 2022.

1 tese eletrônica (149 f.): il. (algumas color.).

Orientador: Fernando França da Cunha.

Tese (doutorado) - Universidade Federal de Viçosa,  
Departamento de Engenharia Agrícola, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.411>

Modo de acesso: World Wide Web.

1. Evapotranspiração - Estimativas. 2. Inteligência artificial.  
3. Irrigação. 4. Hidrologia. I. Cunha, Fernando França da, 1980-.  
II. Universidade Federal de Viçosa. Departamento de Engenharia  
Agrícola. Programa de Pós-Graduação em Engenharia Agrícola.  
III. Título.

CDD 22. ed. 551.572

Bibliotecário(a) responsável: Alice Regina Pinto CRB6 2523

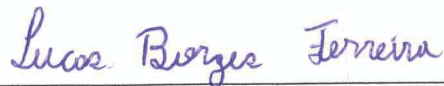
**LUCAS BORGES FERREIRA**

**APPLICATIONS OF MACHINE LEARNING FOR REFERENCE  
EVAPOTRANSPIRATION MODELING**

Thesis submitted to the Agricultural Engineering  
Graduate Program of the Universidade Federal de  
Viçosa in partial fulfillment of the requirements  
for the degree of *Doctor Scientiae*.

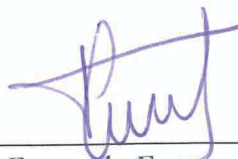
APPROVED: February 10, 2022.

Assent:



---

Lucas Borges Ferreira  
Author



---

Fernando França da Cunha  
Adviser

*To God, Lord of life and science,  
To my beloved mother and grandmother,  
To my wife and son, my gifts from God,  
I dedicate!*

## **ACKNOWLEDGMENTS**

To God, for giving me the grace to walk and get here and to take care of me during every second of my live.

To my mother, Diuza Borges Celestina, and grandmother, Maria Danieta Celestina, who supported me throughout my live, gave me love, taught me honesty, faith and countless other teachings.

To my wife, Anunciene Barbosa Duarte, and my son, Samuel Borges Duarte, for being by my side on good and bad days, for loving me and being gifts from God in my life.

To my advisor, Fernando França da Cunha, for his friendship, patience and all the teachings during the doctorate.

To colleagues from Grupo de Estudos e Soluções para Agricultura Irrigada (GESAI), for the good moments and knowledge exchange.

To the friends from Ministério Universidades Renovadas (MUR), for the wonderful experiences lived together.

To all my friends, for making my days happier.

To the professors and staff of the Department of Agricultural Engineering (DEA).

To the Universidade Federal de Viçosa (UFV) for the opportunity to undertake the Doctor degree.

To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for providing scholarship.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To all who, directly or indirectly, helped me during this journey.

Thank you!

## ABSTRACT

FERREIRA, Lucas Borges, D.Sc., Universidade Federal de Viçosa, February, 2022. **Applications of machine learning for reference evapotranspiration modeling**. Adviser: Fernando França da Cunha.

Reliable estimates of reference evapotranspiration (ET<sub>o</sub>) are of great importance in areas such as irrigation scheduling, hydrological studies, water resources management, among others. However, in scenarios of limited availability of meteorological data, estimating ET<sub>o</sub> becomes challenging. In addition, it is also important to predict future ET<sub>o</sub> values, which can help, for example, in irrigation scheduling. Therefore, in the present thesis, several approaches were studied, which were mainly based on machine learning techniques, aiming to estimate ET<sub>o</sub> under limited availability of meteorological data, in addition to predicting future ET<sub>o</sub> daily values. Strategies for evaluating the performance of alternative ET<sub>o</sub> models when used in irrigation management were also evaluated. It was observed that the application of traditional machine learning techniques and deep learning models showed great potential for modeling ET<sub>o</sub> in the different conditions evaluated. The use of hourly temperature and relative humidity data measured throughout the day, combined with hourly extraterrestrial solar radiation, has shown to be a very promising approach to estimate daily ET<sub>o</sub>. In this approach, 1D convolutional neural networks (1D CNN) had better results than the other models evaluated. Thus, the combination of hourly data and 1D CNN models resulted in markedly superior performances than those observed estimating ET<sub>o</sub> based on daily data (conventional approach). When predicting ET<sub>o</sub> values for the next seven days, in general, the MIMO (multiple input multiple output) prediction strategy was the best alternative, offering good performance and lower computational cost. The deep learning models evaluated performed slightly better than the traditional machine learning models evaluated, and both approaches resulted in better results than using historical monthly means as a prediction of future ET<sub>o</sub> values. By employing machine learning models to estimate ET<sub>o</sub> and some commonly unavailable meteorological data (relative humidity, solar radiation and wind speed), superior performances were observed in relation to the performance obtained with traditional methodologies. The use of multi-task learning to estimate, in a combined way, missing meteorological data and ET<sub>o</sub> resulted in performances similar to those observed considering individual estimations (single-task learning). Finally, it was found that, in addition to using error metrics, such as root mean square error (RMSE), the evaluation of ET<sub>o</sub> models must also consider the behavior of the models

throughout the year. Furthermore, simulating the application of ETo models in irrigation scheduling can provide valuable information for choosing the most appropriate model.

**Keywords:** Deep learning. CNN. Irrigation. LSTM. Time series.

## RESUMO

FERREIRA, Lucas Borges, D.Sc., Universidade Federal de Viçosa, fevereiro de 2022. **Aplicações de aprendizado de máquina para a modelagem de evapotranspiração de referência**. Orientador: Fernando França da Cunha.

Estimar a evapotranspiração de referência (ET<sub>o</sub>) de forma confiável é de grande valia para áreas como manejo da irrigação, estudos hidrológicos, gestão de recursos hídricos, dentre outros. No entanto, em cenários de limitada disponibilidade de dados meteorológicos, estimar a ET<sub>o</sub> torna-se desafiador. Além disso, é também importante prever valores de ET<sub>o</sub> futuros, o que pode auxiliar, por exemplo, na programação das irrigações. Assim, na presente tese, foram estudadas diversas abordagens, baseadas, principalmente, em técnicas de aprendizado de máquina, visando estimar a ET<sub>o</sub> sob limitada disponibilidade de dados meteorológicos, além de prever valores futuros de ET<sub>o</sub> diária. Avaliou-se ainda estratégias para a avaliação do desempenho de modelos alternativos de ET<sub>o</sub> quando empregados no manejo da irrigação. Foi observado que a aplicação de técnicas de aprendizado de máquina tradicionais e modelos de aprendizado profundo apresentou amplo potencial para a modelagem da ET<sub>o</sub> nas diversas condições avaliadas. O uso de dados de temperatura e umidade relativa horários medidos ao longo do dia, combinados com a radiação solar no topo da atmosfera horária, mostrou-se uma abordagem bastante promissora para estimar a ET<sub>o</sub> diária. Nesta abordagem, as redes neurais convolucionais 1D (1D CNN) apresentaram resultados superiores aos demais modelos avaliados. Assim, a combinação de dados horários e modelos 1D CNN resultou em desempenhos destacadamente superiores aos observados estimando a ET<sub>o</sub> com base em dados diários (abordagem convencional). Ao prever a ET<sub>o</sub> dos sete dias futuros, em geral, a estratégia de previsão MIMO (multiple input multiple output) foi a melhor alternativa, oferecendo boa performance e menor custo computacional. Os modelos de aprendizado profundo avaliados apresentaram desempenho ligeiramente superior do que os modelos de aprendizado de máquina tradicionais avaliados e, ambas as abordagens, resultaram em performances melhores do que as obtidas com o uso de médias mensais históricas como previsão de valores futuros de ET<sub>o</sub>. Ao empregar modelos de aprendizado de máquina para estimar a ET<sub>o</sub> e alguns dados meteorológicos comumente indisponíveis (umidade relativa, radiação solar e velocidade do vento), foram observados desempenhos superiores do que os obtidos com metodologias tradicionais. O uso de multi-task learning para estimar, de forma combinada, os dados meteorológicos faltantes e a ET<sub>o</sub> resultou em performances similares às observadas realizando



as estimativas de forma individual (single-task learning). Por fim, verificou-se que além de utilizar métricas de erro, como a raiz do erro quadrático médio (RMSE), a avaliação de modelos de ETo deve também considerar o comportamento dos modelos ao longo do ano. Ademais, simular a aplicação de modelos ETo no manejo da irrigação pode fornecer informações valiosas para a escolha do modelo mais adequado.

**Palavras-chave:** Aprendizado profundo. CNN. Irrigação. LSTM. Série temporal.

## TABLE OF CONTENTS

<b>General introduction</b> .....	10
<b>References</b> .....	12
<b>Scientific articles</b> .....	14
<b>Article 1: New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning</b> .....	14
<b>1 Introduction</b> .....	14
<b>2 Materials and methods</b> .....	16
<b>3 Results and discussion</b> .....	24
<b>4 Conclusions</b> .....	38
<b>References</b> .....	39
<b>Article 2: Multi-step ahead forecasting of daily reference evapotranspiration using deep learning</b> .....	45
<b>1 Introduction</b> .....	45
<b>2 Materials and methods</b> .....	48
<b>3 Results and discussion</b> .....	59
<b>4 Conclusions</b> .....	74
<b>References</b> .....	76
<b>Article 3: Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil</b> .....	80
<b>1 Introduction</b> .....	80
<b>2 Materials and methods</b> .....	83
<b>3 Results and discussion</b> .....	94
<b>4 Conclusions</b> .....	115
<b>References</b> .....	117
<b>Appendix A. Supplementary data</b> .....	122
<b>Article 4: Selecting models for the estimation of reference evapotranspiration for irrigation scheduling purposes</b> .....	125
<b>1 Introduction</b> .....	125
<b>2 Materials and methods</b> .....	127
<b>3 Results and discussion</b> .....	132
<b>4 Conclusions</b> .....	146
<b>References</b> .....	147
<b>General conclusions</b> .....	149

## General introduction

Evapotranspiration (ET) is constituted by the sum of the water transferred to the atmosphere through the processes of evaporation and plant transpiration. Knowing ET is of great importance in several areas, such as irrigation scheduling and design, water resources management, hydrological studies and agrometeorology. In irrigation scheduling, ET is used as a basis for determining the amount of water that should be applied, via irrigation, to ensure a proper crop development.

There are several alternatives to compute ET, such as the use of lysimeters, Bowen ratio method, eddy covariance, among others. However, ET is mostly estimated based on reference evapotranspiration (ET<sub>o</sub>), crop coefficient (K<sub>c</sub>) and other adjustment coefficients (Pereira et al., 2015). Therefore, it is extremely important to accurately determine ET<sub>o</sub>, which contributes to obtain good ET estimates.

ET<sub>o</sub> can be estimated based on meteorological data and mathematical equations. Among the methodologies available for this purpose, the FAO-56 Penman-Monteith equation (FAO56-PM) is one of the best approaches, given its recognized reliability (Pereira et al., 2015). However, to use this equation, air temperature, relative humidity, wind speed and solar radiation data are needed. Thus, given the common partial or total unavailability of these data in different parts of Brazil and the world, the use of the FAO56-PM equation presents certain limitations (Hadria et al., 2021; Paredes and Pereira, 2019).

In order to estimate ET<sub>o</sub> under conditions of limited availability of meteorological data, several empirical equations with reduced data requirements were developed. The use of models based on air temperature is especially important, given the greater availability of this meteorological variable (Mattar et al., 2016; Paredes and Pereira, 2019). In addition, in relation to the other meteorological variables mentioned above, temperature can be measured using less expensive sensors. The use of relative humidity, in addition to temperature data, can contribute to obtaining better ET<sub>o</sub> estimates at a low additional cost (Ferreira et al., 2019; Exner-Kitttridge, 2012; Valiantzas, 2012; Valiantzas, 2018).

In some scenarios, even with the unavailability of solar radiation and wind speed data, which normally require higher cost sensors to be measured, temperature and relative humidity data measured at a high temporal frequency may be available. In these cases, the potential of the data measured in a high frequency, such as on a hourly scale, can be explored to estimate daily ET<sub>o</sub>, which is commonly computed for irrigation management based on data measured on a daily scale. It is believed that hourly meteorological data measured throughout the day may contain information capable of helping models to estimate daily ET<sub>o</sub> with better accuracy,

even using only on air temperature data as a basis. This hypothesis is raised based on the relationship between the variation of air temperature throughout the day and the values assumed by other meteorological variables that influence ETo and that are not being used as input for a prediction model. In this way, the behavior of air temperature throughout the day could serve as information capable of partially supplying the lack of some meteorological data not available during the determination of ETo.

In addition to the traditional empirical equations used to estimate ETo with lower meteorological data requirements, machine learning models have been used for this purpose (Ferreira et al., 2019; Huang et al., 2019; Kisi and Alizamir, 2018; Yu et al., 2020). In general, such models have a high capacity to capture complex patterns between input variables and ETo. Thus, machine learning models generally perform better when compared to traditional empirical equations.

In addition to developing models to estimate ETo under limited data availability, it is important to create models to predict future ETo values. For example, knowing ahead ETo values can help in planning future irrigations. In this context, machine learning models can also be explored (Ashrafzadeh et al., 2020; Karbasi, 2018).

When addressing time-series predictions, such as ETo forecasting, or modeling sequential data, such as using hourly data to estimate daily ETo, deep learning models can also be employed (Canizo et al., 2019; Gao et al., 2019; Ng et al., 2019). Deep learning is a sub-area of machine learning. This field has gained great attention in recent years, being successfully used in the modeling of time series and other sequential data, in addition to having important applications in areas such as computer vision and natural language processing.

In addition to the use of good modeling techniques, it is of fundamental importance that the models developed to estimate ETo are well evaluated. This step is responsible for indicating whether a particular model presents satisfactory performance. In addition to the use of error metrics, such as root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination ( $R^2$ ), the simulation of the use of ETo models in irrigation scheduling can provide information capable of supporting the choice of the most appropriate model for irrigation scheduling purposes.

In view of the foregoing, the objective was to explore the potential of machine learning and deep learning techniques to estimate ETo in different scenarios of data availability, as well as to estimate future ETo values and to propose a methodology for evaluating alternative ETo models when used in irrigation scheduling. Therefore, this thesis was composed of four scientific articles, with the following titles: (1) New approach to estimate daily reference

evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning; (2) Multi-step ahead forecasting of daily reference evapotranspiration using deep learning; (3) Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil; and (4) Selecting models for the estimation of reference evapotranspiration for irrigation scheduling purposes.

## References

- Ashrafzadeh, A., Kişi, O., Aghelpour, P., Biazar, S.M., Masouleh, M.A., 2020. Comparative Study of Time Series Models, Support Vector Machines, and GMDH in Forecasting Long-Term Evapotranspiration Rates in Northern Iran. *J. Irrig. Drain. Eng.* 146. doi:10.1061/(ASCE)IR.1943-4774.0001471
- Canizo, M., Triguero, I., Conde, A., Onieva, E., 2019. Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study. *Neurocomputing.* 363, 246-260. doi:10.1016/j.neucom.2019.07.034
- Exner-Kittridge, M., 2012. Closure to “Case Study on the Accuracy and Cost/Effectiveness in Simulating Reference Evapotranspiration in West-Central Florida” by Michael Grant Exner-Kittridge and Mark Cable Rains
- Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. *J. Hydrol.* 572, 556–570. doi:10.1016/j.jhydrol.2019.03.028
- Gao, S., Wang, X., Miao, X., Su, C., Li, Y., 2019. ASM1D-GAN: An Intelligent Fault Diagnosis Method Based on Assembled 1D Convolutional Neural Network and Generative Adversarial Networks. *J. Signal Process. Syst.* 91, 1237-1247. doi:10.1007/s11265-019-01463-8
- Hadria, R., Benabdelouhab, T., Lionboui, H., Salhi, A., 2021. Comparative assessment of different reference evapotranspiration models towards a fit calibration for arid and semi-arid areas. *J. Arid Environ.* 184, 104318. doi:10.1016/j.jaridenv.2020.104318
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* 574, 1029–1041. doi:10.1016/j.jhydrol.2019.04.085
- Karbasi, M., 2018. Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet- Gaussian Process Regression Model. *Water Resour. Manag.* 32, 1035–1052. doi:10.1007/s11269-017-1853-9

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. doi:10.1016/j.agrformet.2018.08.007

Mattar, M.A., Alazba, A.A., Alblewi, B., Gharabaghi, B., Yassin, M.A., 2016. Evaluating and Calibrating Reference Evapotranspiration Models Using Water Balance under Hyper-Arid Environment. *Water Resour. Manag.* 30, 3745–3767. doi:10.1007/s11269-016-1382-y

Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*. 352, 251-267. doi: 10.1016/j.geoderma.2019.06.016

Paredes, P., Pereira, L.S., 2019. Computing FAO56 reference grass evapotranspiration PM-ET<sub>0</sub> from temperature with focus on solar radiation. *Agric. Water Manag.* 215, 86–102. doi:10.1016/j.agwat.2018.12.014

Pereira, L.S., Allen, R.G., Smith, M., Raes, D., 2015. Crop evapotranspiration estimation with FAO56: Past and future. *Agric. Water Manag.* doi:10.1016/j.agwat.2014.07.031

Valiantzas, J.D., 2012. Discussion of “Case Study on the Accuracy and Cost/Effectiveness in Simulating Reference Evapotranspiration in West-Central Florida” by Michael Grant Exner-Kittridge and Mark Cable Rains. *J. Hydrol. Eng.* 17, 224–225. doi:10.1061/(asce)he.1943-5584.0000394

Valiantzas, J.D., 2018. Temperature-and humidity-based simplified Penman’s ET<sub>0</sub> formulae. Comparisons with temperature-based Hargreaves-Samani and other methodologies. *Agric. Water Manag.* 208, 326–334. doi:10.1016/j.agwat.2018.06.028

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. doi:10.1016/j.compag.2020.105653

## Scientific articles

### Article 1: New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning

**Abstract:** Computation of reference evapotranspiration (ET<sub>o</sub>) poses a challenge under limited meteorological data availability. However, even in this case, hourly data may be available since low-cost sensors can report hourly measurements. This study evaluates, for the first time, in regional and local scenarios, the use of limited hourly meteorological data (temperature and relative humidity or only temperature) to estimate daily ET<sub>o</sub>, employing RF, XGBoost, ANN and CNN. The following options were evaluated: (i) use of daily input data (conventional approach); (ii) use of hourly data measured during a 24 h period + hourly extraterrestrial radiation (Ra) to estimate daily ET<sub>o</sub> directly; and (iii) the same configuration of the last option, but with daily Ra instead of hourly Ra. All options used Ra. To develop and evaluate the models, two daily ET<sub>o</sub> targets were considered: ET<sub>o,d</sub> (computed using the daily version of the ASCE-PM equation) and ET<sub>o,soh</sub> (computed by summing hourly ET<sub>o</sub> obtained with the ASCE-PM equation). Data from 53 weather stations located in the state of Minas Gerais, Brazil, were used. For all models, the best results were found using hourly data to estimate daily ET<sub>o</sub> directly. CNN models developed with 24 h hourly data + hourly Ra offered the best performance in all cases. In relation to the best models developed with daily data, RMSE reduced by up to 28.2% (0.71 to 0.51) and NSE and R<sup>2</sup> increased by up to 21.7 (0.69 to 0.84) and 11.4% (0.79 to 0.88), respectively, in regional scenario. In local scenario, RMSE reduced by up to 22.4% (0.58 to 0.45) and NSE and R<sup>2</sup> increased by up to 10.1 (0.79 to 0.87) and 11.3% (0.80 to 0.89), respectively.

**Keywords:** 1D CNN, convolutional neural network, irrigation scheduling, neural network, random forest

## 1 Introduction

Reliable estimates of reference evapotranspiration (ET<sub>o</sub>) are crucial in tasks such as irrigation scheduling, water resource management and hydrological studies. Given the complexity involved with the direct measurement of this parameter, it is commonly estimated based on meteorological data (Pereira et al., 2015). To accomplish this, the FAO-56 Penman-Monteith (FAO56-PM) equation (Allen et al., 1998) or the ASCE (American Society of Civil Engineers) Penman-Monteith (ASCE-PM) equation (ASCE-EWRI, 2005) can be used. These equations, although well accepted in practical applications and academic research, require meteorological data (temperature, relative humidity, solar radiation and wind speed) that may

be partially unavailable in some locations. Therefore, several studies have investigated alternative equations, such as the well-known Hargreaves-Samani equation (Hargreaves and Samani, 1985), to estimate ETo with reduced data requirement (Ahooghalandari et al., 2016; Almorox et al., 2018; Paredes and Pereira, 2019; Valiantzas, 2018; Zanetti et al., 2019).

In the last years, machine learning models have been successfully used to estimate ETo with fewer meteorological data. These models are capable of capturing complex relationships between input and output data, which makes them powerful tools in ETo modeling. Various models have been assessed, such as artificial neural network (ANN) (Ferreira et al., 2019; Kumar et al., 2011; Nourani et al., 2019; Wu and Fan, 2019), support vector machine (SVM) (Ferreira et al., 2019; Mehdizadeh et al., 2017; Nourani et al., 2019), adaptive neuro-fuzzy inference system (ANFIS) (Nourani et al., 2019; Wu and Fan, 2019), extreme learning machine (ELM) (Abdullah et al., 2015; Fan et al., 2018), multivariate adaptive regression splines (MARS) (Mehdizadeh et al., 2017; Wu and Fan, 2019), random forest (RF) (Fan et al., 2019; Feng et al., 2017; Wang et al., 2019), extreme gradient boosting (XGBoost) (Fan et al., 2018; Wu and Fan, 2019) and light gradient boosting machine (LightGBM) (Fan et al., 2019). In general, machine learning models have outperformed conventional equations for estimation of ETo, reaching higher performances for the same data requirement (Fan et al., 2019; Ferreira et al., 2019; Kiafar et al., 2017; Kumar et al., 2011; Mehdizadeh et al., 2017; Reis et al., 2019).

In the standard approach to develop a model to estimate daily ETo, daily meteorological data are used as input. However, in some cases, even with limited availability of meteorological data, hourly data may be available. For instance, a simplified automatic weather station with only an air temperature sensor can report hourly measurements. In this situation, only daily maximum and minimum temperatures would be used when considering the standard approach, losing any possible extra information contained in hourly data. To obtain possible benefits in using limited hourly data to estimate daily ETo, models can be developed to estimate daily ETo directly using all the hourly data measured during a 24 h time period. However, according to our knowledge, so far no study investigated the use of limited hourly data to estimate daily ETo directly.

It should be highlighted that models that can capture additional information from hourly data may achieve higher accuracy in estimation of daily ETo. It is especially important for temperature-based models due to the following reasons: (i) temperature-based models commonly do not exhibit high performances; (ii) temperature sensors are generally cheaper than sensors to measure relative humidity, wind speed and solar radiation; (iii) advances in estimation of ETo using only air temperature can enable the development of low-cost systems



for irrigation scheduling. It is also worth mentioning that the inclusion of relative humidity in addition to temperature in ETo models can contribute to achieve higher performance at a low additional cost (Exner-Kittridge, 2012; Valiantzas, 2012; Valiantzas, 2018).

When using hourly data obtained during a 24 h time period, it is possible to employ models that can handle sequential data. In this sense, convolutional neural network (CNN) can be used. CNN is a deep learning model that has gained a lot of attention in image recognition field, including applications in agricultural area (Dyrmann et al., 2016; Ji et al., 2018; Kamilaris and Prenafeta-Boldú, 2018a; Kamilaris and Prenafeta-Boldú, 2018b; LeCun et al., 2015). In the case of sequential data or time series, like hourly meteorological data, one-dimensional convolutional neural network (1D CNN), an especial type of CNN, can be used (Abdeljaber et al., 2018; Canizo et al., 2019; Gao et al., 2019; Li et al., 2017; Liu et al., 2018; Ng et al., 2019). In hydrology/climatology field, studies using 1D CNN are still very scarce (Haidar and Verma, 2018). Generally, CNN has outperformed traditional machine learning models in many studies, reaching state-of-the-art performances. Despite its high capabilities, deep learning models, such as CNN, have not been very explored in hydrological sciences, as reported for Shen (2018) in an extensive review study. In estimation of ETo, according to our knowledge, so far CNN was not used.

Considering the importance of reliable estimates of ETo using reduced datasets, the objective of the present study is to assess, for the first time, in regional and local scenarios, the use of limited hourly meteorological data (temperature and relative humidity or only temperature) to estimate daily ETo directly, employing RF, XGBoost, ANN and CNN.

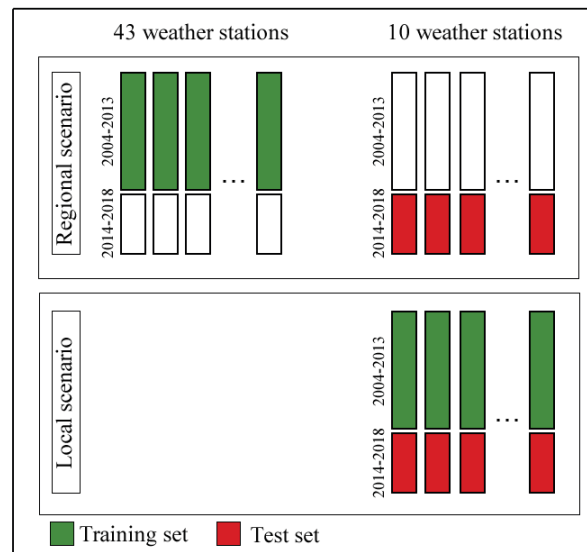
## **2 Materials and methods**

### **2.1 Database, study sites and data management**

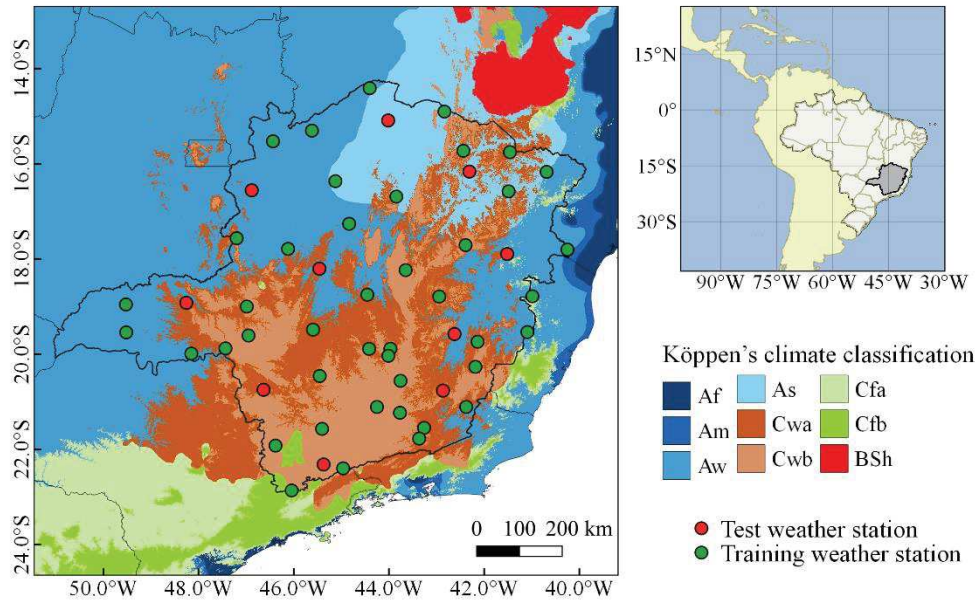
Hourly data from 53 automatic weather stations of the Brazilian National Institute of Meteorology (INMET) were used. All the stations are located in the state of Minas Gerais, Brazil. The data length varied according to the stations, given their different epochs of creation, with a mean length of  $11.7 \pm 1.34$  years. All the stations have data up to the year 2018. Maximum and minimum air temperature, maximum and minimum relative humidity, solar radiation and wind speed were collected. Wind speed, measured at 10 m height, was converted to 2 m, according to Allen et al. (1998). Days with missing or faulty data were removed. Faulty data were identified in daily scale when minimum temperature was higher than maximum temperature; relative humidity out of the range 0-100%; minimum relative humidity higher than

maximum relative humidity; negative wind speed; negative solar radiation; or solar radiation higher than extraterrestrial radiation.

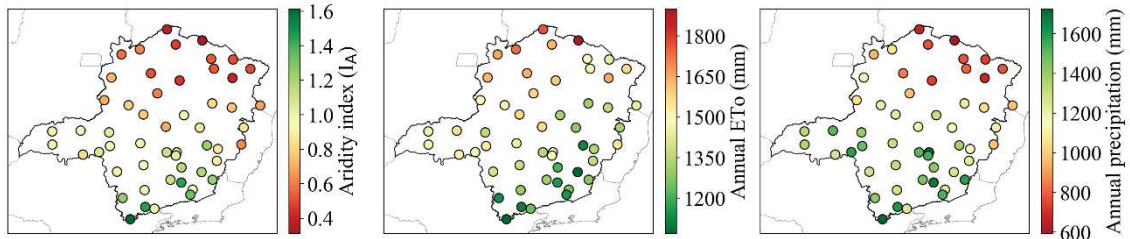
To perform the analysis, two scenarios were considered: (i) regional, models were trained with pooled data from 43 weather stations and evaluated in the 10 remaining stations; and (ii) local, models were trained and evaluated using individual data from the 10 weather stations used for evaluation in the previous scenario. The 10 stations mentioned were selected in order to represent different climatic conditions. In both regional and local scenarios, data up to 2013 were used in the training process of the models and data from 2014 to 2018 were used to test them (Fig. 1). The weather stations split, as well as their locations and climate classification (Alvares et al., 2013) can be seen in Fig. 2. Aridity index ( $I_A$ ) (UNEP, 1997) (ratio between annual precipitation to annual reference evapotranspiration), annual reference evapotranspiration and annual precipitation are presented in Fig. 3. It can be noted that the study area covers a wide variety of climatic conditions, with  $I_A$  ranging from 0.31 to 1.61.



**Fig. 1.** Flowchart of the data split scheme used.



**Fig. 2.** Location, climate classification and training/test split of the weather stations used.



**Fig. 3.** Aridity index ( $I_A$ ), annual reference evapotranspiration and annual precipitation for the weather stations used (2004-2018).

## 2.2 Input data for the machine learning models

To explore the potential of hourly limited meteorological data to estimate daily ETo, the following approaches were employed to develop the machine learning models: (i) use of daily data to estimate daily ETo (conventional approach); (ii) use of hourly data measured during a 24 h period and hourly extraterrestrial radiation ( $R_a$ ) to estimate daily ETo directly; and (iii) use of hourly data measured during a 24 h period and daily  $R_a$  to estimate daily ETo directly. Approach i also used  $R_a$ , in daily scale. The variation in  $R_a$  scale (approaches ii and iii) was done in order to verify the effect of  $R_a$  scale in the estimation of ETo.

Two meteorological data availability options were assessed: temperature-based models and temperature- and relative humidity-based models. In the first option, maximum temperature, minimum temperature and  $R_a$  were considered. In the second option, in addition to the data of the first option, maximum and minimum relative humidity were used.

### 2.3 Targets for development and evaluation of the machine learning models

When developing models for estimation of ETo, it is a common practice to use ETo estimated by the FAO56-PM equation or ASCE-PM equation, which is used in this study, as target. For that, one can estimate daily ETo using the daily version of Penman-Monteith equation, here called ETo<sub>d</sub>, or summing hourly ETo obtained with the hourly version of Penman-Monteith equation, here called ETo<sub>soh</sub>. The last option is generally considered more accurate under certain scenarios (Allen et al., 1998; Pereira et al., 2015; Althoff et al., 2019). It is expected that when using ETo<sub>soh</sub> as target, machine learning models developed with hourly input data will perform better than their versions developed with daily data since hourly data were used to calculate the target. Therefore, when considering ETo<sub>d</sub> as additional target, if hourly data also provide performance increments for this target, this gain can be associated with their capability to better capture climate patterns related to ETo, which may not be as clear as when using only ETo<sub>soh</sub> as target. Thus, in this study, the two types of daily ETo targets (i.e., ETo<sub>d</sub> and ETo<sub>soh</sub>) were employed to develop and evaluate the machine learning models for direct estimation of daily ETo (cases i, ii and iii).

To compute ETo in both hourly and daily scale, the ASCE-PM equation (Allen et al., 2006) (Eq. 1) was employed. ASCE-PM differs from FAO-PM in the surface resistance values adopted for hourly time step. However, Allen et al. (2006) recommended to use FAO-PM with the same surface resistance values used in ASCE-PM. The calculation of all parameters necessary to use ASCE-PM was done based on FAO-56 paper.

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

where ETo is the reference evapotranspiration (mm d<sup>-1</sup> or mm h<sup>-1</sup>), R<sub>n</sub> is the net solar radiation (MJ m<sup>-2</sup> d<sup>-1</sup> or MJ m<sup>-2</sup> h<sup>-1</sup>), G is the soil heat flux density (MJ m<sup>-2</sup> d<sup>-1</sup> or MJ m<sup>-2</sup> h<sup>-1</sup>) (considered equal to 0 in daily time step and equal to 0.1 R<sub>n</sub> and 0.5 R<sub>n</sub> for hourly time step during daytime (defined as when R<sub>n</sub> > 0) and nighttime, respectively), T is the mean daily or hourly air temperature (°C), u<sub>2</sub> is the mean daily or hourly wind speed at a 2 m height (m s<sup>-1</sup>), e<sub>s</sub> is the saturation vapour pressure (kPa), e<sub>a</sub> is the actual vapour pressure (kPa) (obtained using maximum and minimum relative humidity for daily time step and mean relative humidity for hourly time step), Δ is the slope of the saturation vapour pressure function (kPa °C<sup>-1</sup>), and γ is the psychrometric constant (kPa °C<sup>-1</sup>). C<sub>n</sub> and C<sub>d</sub> are constants that depend on calculation time step, reference type and time of day. For grass ETo, C<sub>n</sub> is equal to 900 K mm s<sup>3</sup> mg<sup>-1</sup> d<sup>-1</sup> for daily time step and 37 K mm s<sup>3</sup> mg<sup>-1</sup> h<sup>-1</sup> for hourly time step; and C<sub>d</sub> is equal to 0.34 s m<sup>-1</sup> for

daily time step and  $0.24$  and  $0.96 \text{ s m}^{-1}$  for hourly time step during daytime ( $R_n > 0$ ) and nighttime, respectively.

For  $R_n$  computation, the ratio between solar radiation and clear sky solar radiation ( $R_s/R_{so}$ ), which is used to represent cloud cover, was limited to  $\leq 1.0$ . As it is not possible to compute this ratio during nighttime (in this case, defined during computations as when  $R_a$  is equal to 0),  $R_s/R_{so}$  calculated between 2-3 hours before sunset was adopted. As nighttime  $R_s/R_{so}$  (2-3 hours before sunset) from a given day is used in the subsequent nighttime period of the next day, it is not possible to obtain  $R_s/R_{so}$  when the previous day is missing. In this study, a default value (0.6) was used in this case. This value was selected based on recommendations of Allen et al. (1998), which proposes 0.4 to 0.6 in humid and subhumid climates and 0.7 to 0.8 in arid and semiarid climates.

## **2.4 Machine learning models**

All machine learning models were implemented using the Python programming language, with the aid of the following libraries: Scikit-learn, XGBoost, TensorFlow and Keras. Computations were performed using a virtual machine from Google Cloud Platform. To train CNN models, a virtual machine with a graphics processing unit (GPU) was used. The models hyperparameters were defined by grid search. To accomplish this, the data set available to train the models was splitted into two subsets, the first one, named training subset (70% randomly chosen), was used to train the models, and the second one (30%) was used as a validation subset for hyperparameters tuning. Therefore, the hyperparameters values that minimized the error in the validation subset were chosen. Finally, the models were tested with data from the test set.

### **2.4.1 Random forest (RF)**

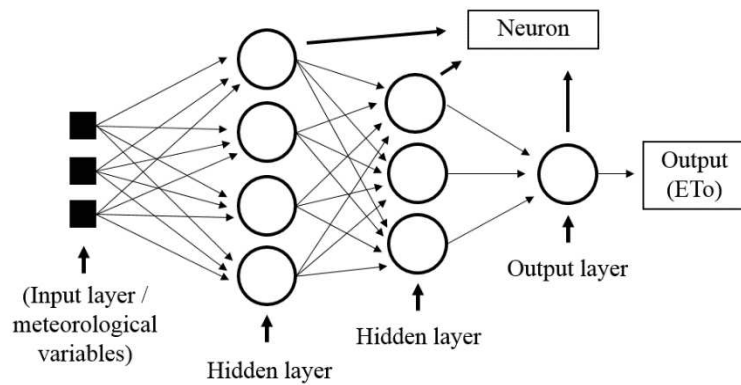
Random forest (RF) is a decision tree-based algorithm proposed by Breiman (2001). This model combines several decision trees fitted in different subsets of the training data. Each tree is considered a weak learner, however, the combination of trees (ensemble) results in a single model with high predictive power (Huang et al., 2019). RF also has the advantage of estimating the importance of each of the input variables. Regarding the training process, RF generally requires less adjustment in hyperparameters. The main hyperparameters to be adjusted, which were adjusted in this study, are the number of trees ( $n\_estimators$ ) and the number of features considered for splitting at each leaf node ( $max\_features$ ). More information about RF can be seen in Tyrallis et al. (2019). The following values were tested for the hyperparameters mentioned above, respectively: (100, 200, 400 and 500 trees), and (all features, the square root of the total number of features and one third of the features).

### **2.4.2 Extreme gradient boosting (XGBoost)**

XGBoost was recently proposed (Chen and Guestrin, 2016) and has gained wide attention in machine learning competitions (Adam-Bourdarios et al., 2015). This algorithm, like RF, is based on decision trees, differing in the way the tree ensemble is constructed. XGBoost uses previous ideas from gradient boosting, creating each tree based on information from previously created trees. XGBoost is also able to estimate the importance of input variables and presents a higher computational efficiency and better ability to deal with overfitting (Fan et al., 2018). More information can be obtained in Chen and Guestrin (2016). In hyperparameters tuning, the number of trees ( $n\_estimators$ ) and maximum tree depth ( $max\_depth$ ) were adjusted. The number of trees values tested were 100, 200, 400 and 500; and the maximum tree depth values tested were 3, 5, 7, 9 and 11.

### 2.4.3 Artificial neural network (ANN)

ANN of the feed-forward multilayer perceptron (MLP) type was used. It is based on a network of artificial neurons connected by weights, which are organized in layers. Typically, an ANN is composed of an input layer, hidden layers and output layer. Each layer can have a different number of neurons. The input layer has the same size as the number of input variables used, the number of neurons in hidden layers is defined by trial and error and the number of neurons in the output layer depends on the problem, however, for regression tasks, one neuron is normally used. An ANN with three input variables, two hidden layers with four and three neurons, and one neuron in the output layer is exemplified in Fig. 4. More information can be obtained in Ferreira et al. (2019).



**Fig. 4.** Artificial neural network architecture example.

In this study, the ANN architectures were defined by trial and error. The following combinations were evaluated: one and two hidden layers with 10, 15 and 20 neurons in each layer for cases that the number of input variables was less than or equal to five (i.e., models trained with daily data); and one and two hidden layers with 40, 50 and 60 neurons in each layer



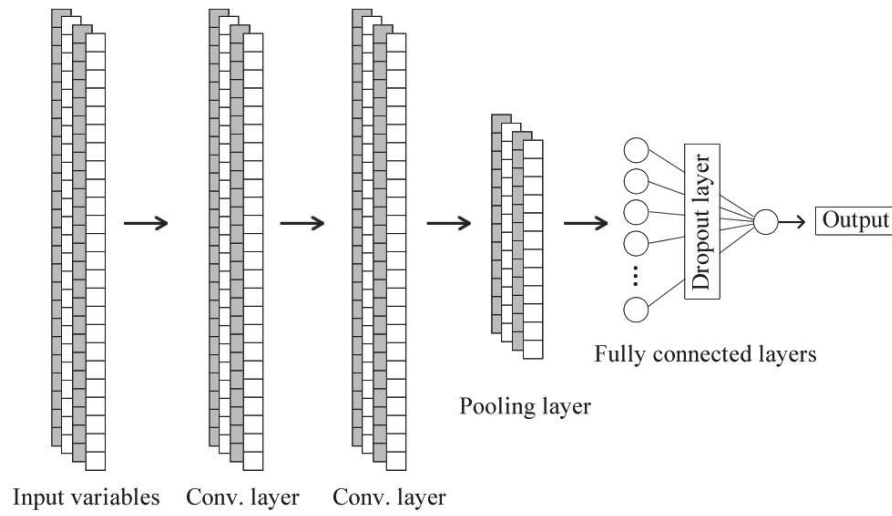
for cases that the number of input variables was greater than 5 (i.e., models trained using 24 h hourly data). Hyperbolic tangent was used as activation function in hidden layers and linear function was used in the output layer. To train the ANN models, the Adam algorithm (Kingma and Ba, 2014) was used. Learning rate was set to 0.001, the number of training epochs was defined using early stopping with maximum training epochs equal to 300 and patience equal to 20 epochs, and batch size was set to 512.

#### **2.4.4 Convolutional neural network (CNN)**

CNN is machine learning model from the deep learning subfield. It is a type of neural network that differs from conventional ANN (i.e., MLP) mainly because it has convolution layers or convolutional filters (Goodfellow et al., 2016). In these layers, an automatic feature extraction process is performed, in which the features of the input data that are really important to map the relationship between the input variables and the desired output are extracted. This ability makes CNN capable of handling raw data without the need for preprocessing or manual feature extraction. Besides convolutional layers, CNN typically has pooling layers, which are used to reduce data dimensionality. More detailed information on CNN can be found in LeCun et al. (2015) and Goodfellow et al. (2016).

CNN is generally applied for image processing. Thus, the convolutional filters used have two dimensions (2D CNN), like an image. However, for the analysis of sequential data or time series, one-dimensional convolutional filters (1D CNN) are used (Li et al., 2017). These filters slide over the input data to capture possible patterns contained in the time dimension, in the case of time series. Therefore, 1D CNN is used in this study. As convolutional layers require input variables with time dimension, this model was used only with the input data combination ii (i.e., 24 h hourly data + hourly Ra).

The CNN architecture used in this study consisted of two convolutional layers, one pooling layer and two fully connected layers with a dropout layer between them (Fig. 5). The number of filters in each convolutional layer, kernel size, number of neurons in the first fully connected layer and dropout rate were defined by trial and error. In each case, the following values, for each hyperparameter, respectively, were tested: (48, 64, 96), (3, 5, 7), (20, 40, 60) and (0, 0.2, 0.4). In convolutional layers, stride was set to 1, padding was set to “same” and the activation function used was relu. In the first fully connected layer, relu was used as activation function and, in the last fully connected layer, linear function was used. To train the models, the Adam algorithm (Kingma and Ba, 2014) was used. Learning rate was set to 0.001, the number of training epochs was defined using early stopping with maximum training epochs equal to 150 and patience equal to 10 epochs, and batch size was set to 512.



**Fig. 5.** Convolutional neural network architecture used in the study.

#### 2.4.5 Data normalization

To avoid convergence problems during the training of the machine learning models, input and output variables were standardized according to Eq. 2. To ensure that the evaluation of the models during validation and test steps can reflect their application in a real case, the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) were computed using only the data of the training subset, not including data from validation subset and test set.

$$x_{ni} = \frac{x_i - \mu}{\sigma} \quad (2)$$

where  $x_{ni}$  is the standardized value,  $x_i$  is the observed value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

#### 2.5 Performance comparison criteria

The models were evaluated for each weather station by means of root mean square error (RMSE), mean bias error (MBE), Nash-Sutcliffe efficiency coefficient (NSE) and coefficient of determination ( $R^2$ ), according to the following equations:

$$RMSE = \sqrt{\frac{1}{n} \sum (P_i - O_i)^2} \quad (3)$$

$$MBE = \frac{1}{n} \sum (P_i - O_i) \quad (4)$$

$$NSE = 1 - \frac{\sum (P_i - O_i)^2}{\sum (\bar{O} - O_i)^2} \quad (5)$$

$$R^2 = \left[ \frac{\sum (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{(\sum (P_i - \bar{P})^2)(\sum (O_i - \bar{O})^2)}} \right]^2 \quad (6)$$



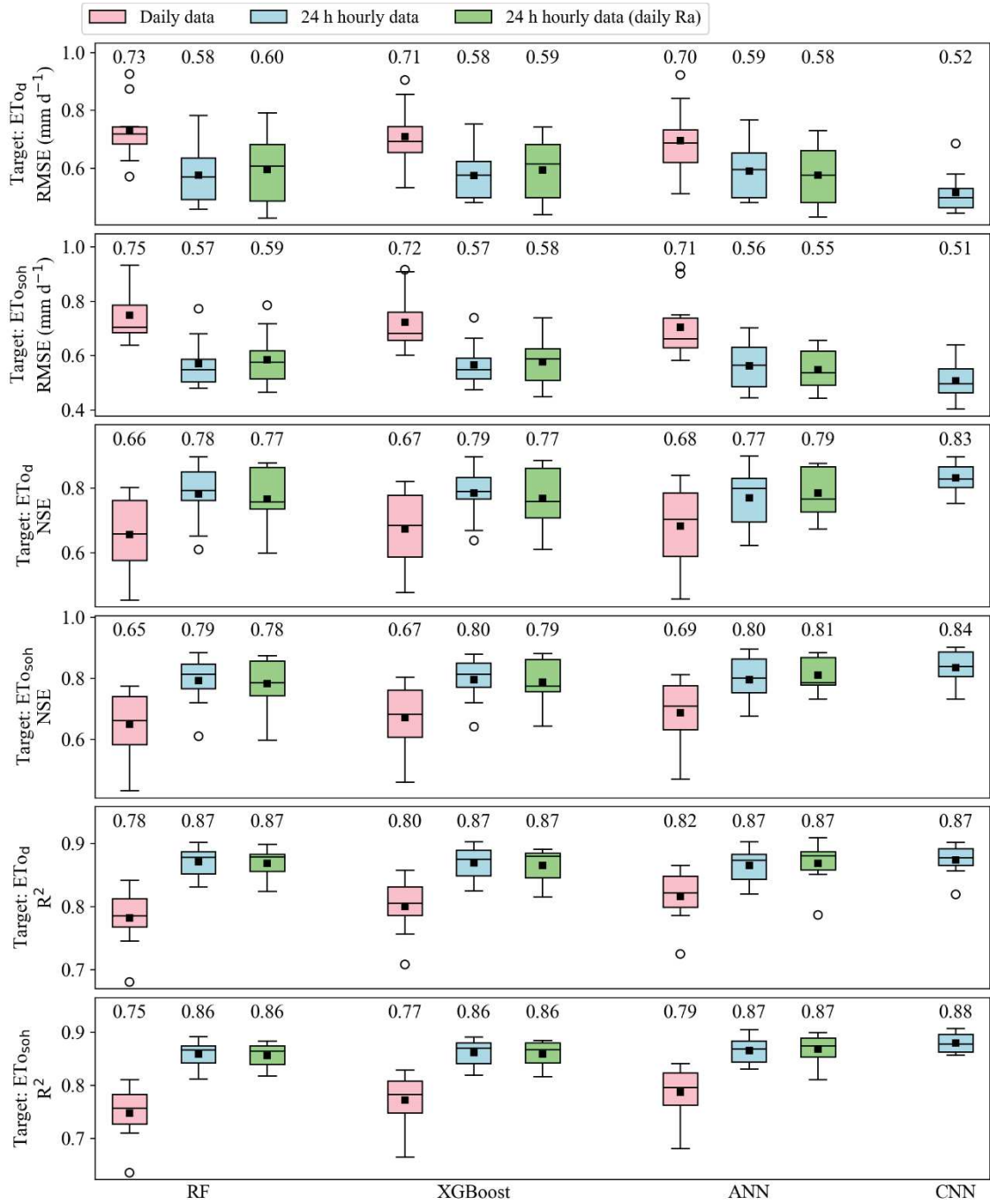
where  $P_i$  is the predicted value ( $\text{mm d}^{-1}$ ),  $O_i$  is the observed value ( $\text{mm d}^{-1}$ ),  $\bar{P}$  is the mean of the predicted values ( $\text{mm d}^{-1}$ ),  $\bar{O}$  is the mean of the observed values ( $\text{mm d}^{-1}$ ), and  $n$  is the number of data pairs.

### **3 Results and discussion**

#### **3.1 Regional scenario**

##### **3.1.1 Temperature-based models**

The results obtained when using  $ET_{od}$  and  $ET_{soh}$  as targets (Fig. 6) generally had the same behavior. All the machine learning models developed with 24 h hourly data using daily or hourly  $R_a$  exhibited better performances than their versions developed with daily input data (conventional approach), showing lower RMSE and higher NSE and  $R^2$  values. The use of hourly  $R_a$  in relation to daily  $R_a$  did not promote significant changes in performance.



**Fig. 6.** Boxplots and mean values of RMSE, NSE and  $R^2$  for regional temperature-based models developed using  $ETo_d$  and  $ETo_{soh}$  as target.

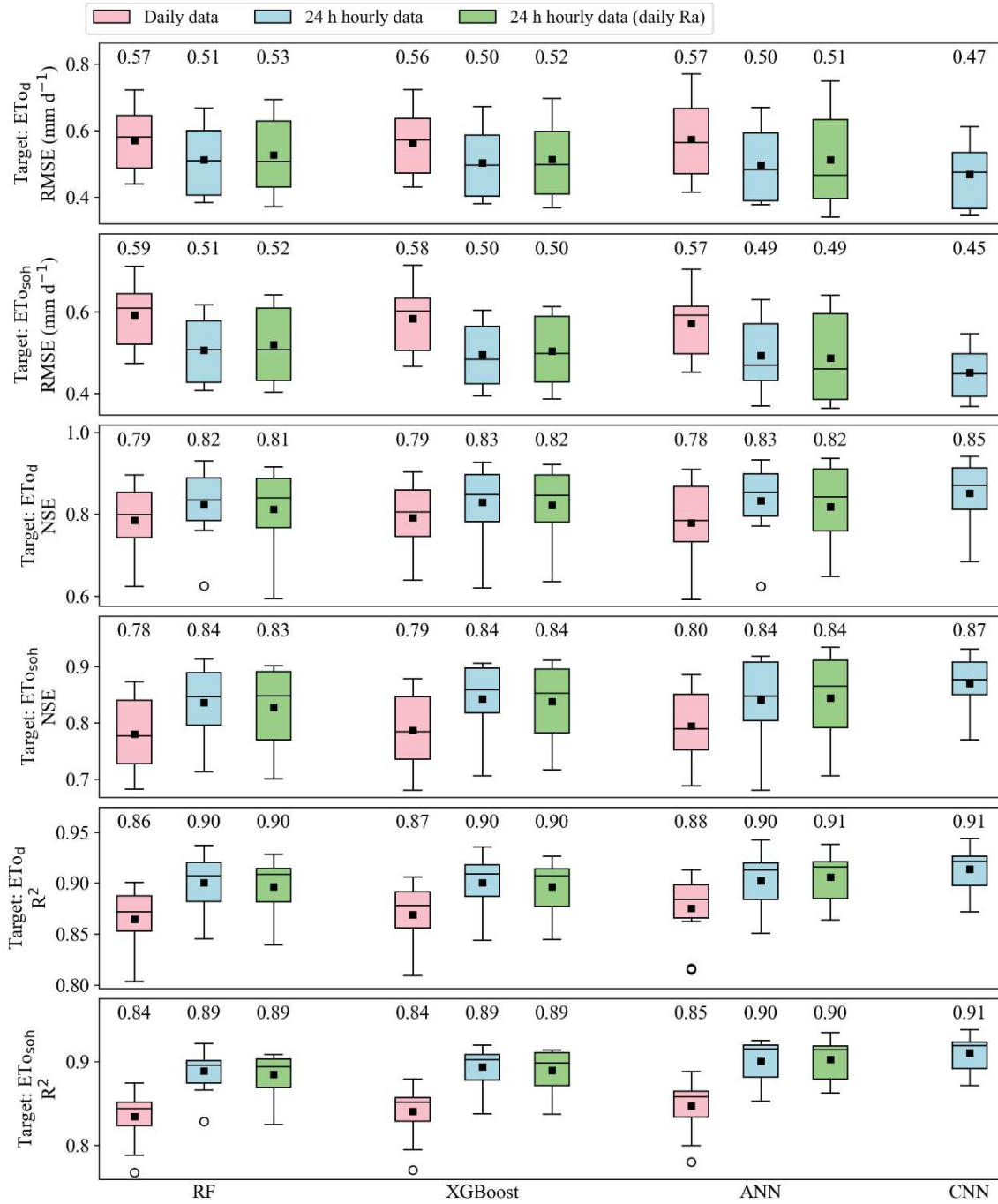
When comparing RF, XGBoost and ANN for all data input combinations evaluated, the best performances were generally obtained for ANN, XGBoost and RF, in this order. However, when 24 h hourly data with hourly Ra were used as input, CNN performed better than all the models, with expressive reductions in RMSE values and increments in NSE values. In relation to ANN, RMSE reduced from 0.59 to 0.52 (11.9%) and from 0.56 to 0.51 (8.9%); and NSE

increased from 0.77 to 0.83 (7.8%) and from 0.80 to 0.84 (5.0%), when considering  $ET_{od}$  and  $ET_{soh}$  as targets, respectively.

Comparing the best results, which were found using CNN (for both  $ET_{od}$  and  $ET_{soh}$  targets) developed using 24 h hourly data and hourly  $R_a$ , with the results found for the best model that used daily data (ANN, for both  $ET_{od}$  and  $ET_{soh}$  targets), RMSE reduced from 0.70 to 0.52 (25.7%) and from 0.71 to 0.51 (28.2%); NSE increased from 0.68 to 0.83 (22.1%) and from 0.69 to 0.84 (21.7%); and  $R^2$  increased from 0.82 to 0.87 (6.1%) and from 0.79 to 0.88 (11.4%), when using  $ET_{od}$  and  $ET_{soh}$  as targets, respectively.

### **3.1.2 Temperature- and relative humidity-based models**

For the regional temperature- and relative humidity-based models, the results obtained for  $ET_{od}$  and  $ET_{soh}$  targets (Fig. 7) had similar behaviors. Again, the best results were found for the models that used 24 h hourly data as input. Comparing the machine learning models, ANN was slightly better than RF and XGBoost. However, CNN exhibited the best performance. In relation to the ANN models developed with 24 h hourly data and hourly  $R_a$ , RMSE reduced from 0.50 to 0.47 (6.0%) and from 0.49 to 0.45 (8.2%); NSE increased from 0.83 to 0.85 (2.4%) and from 0.84 to 0.87 (3.6%); and  $R^2$  increased from 0.90 to 0.91 (1.1%) and from 0.90 to 0.91 (1.1%), when considering  $ET_{od}$  and  $ET_{soh}$  as targets, respectively.



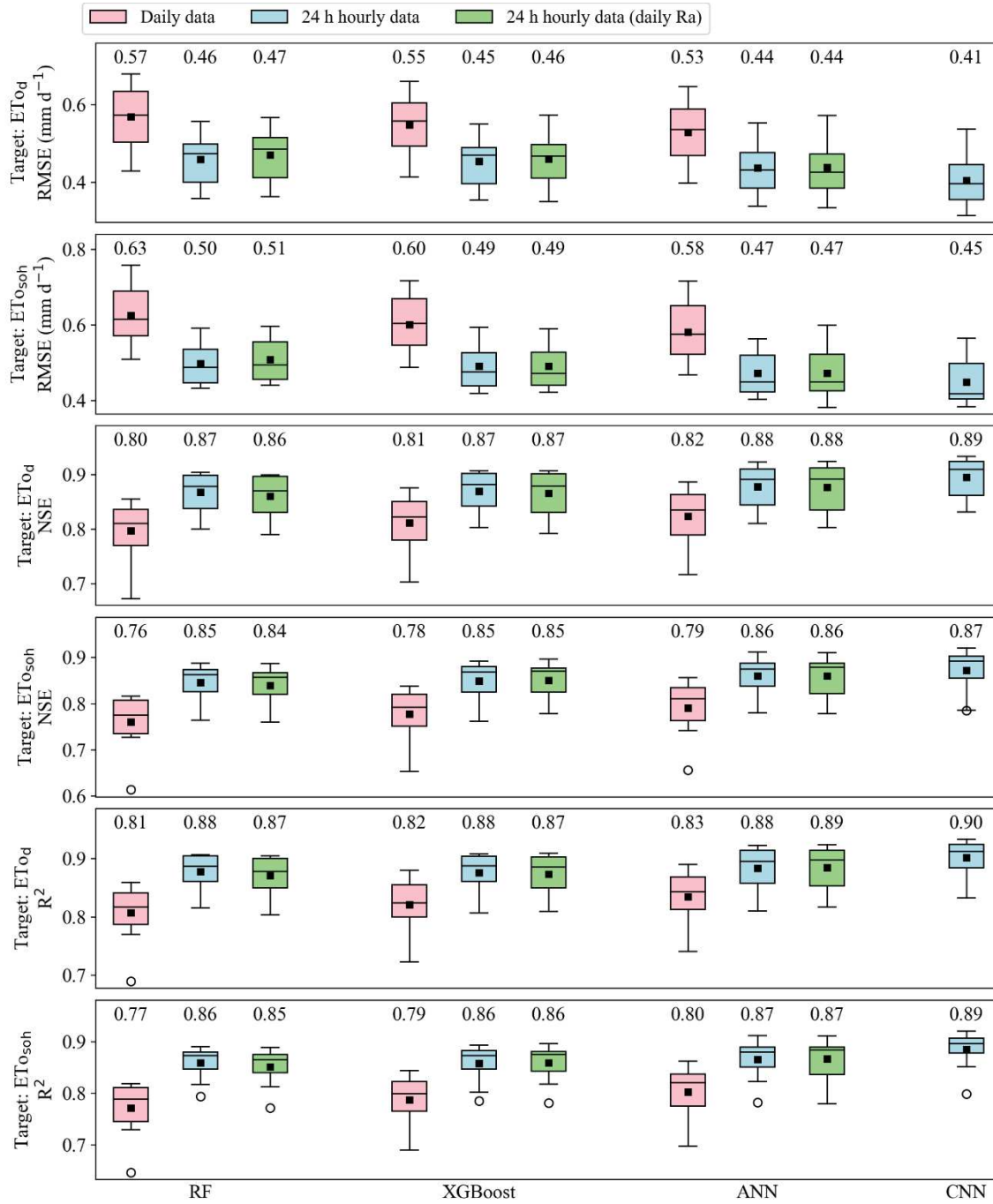
**Fig. 7.** Boxplots and mean values of RMSE, NSE and  $R^2$  for regional temperature- and relative humidity-based models developed using  $ET_{od}$  as target.

Comparing the best results, which were found using CNN (for both  $ET_{od}$  and  $ET_{osoh}$  targets) developed using 24 h hourly data and hourly Ra, with the results found for the best models that used daily data (XGBoost and ANN for  $ET_{od}$  and  $ET_{osoh}$  targets, respectively), RMSE reduced from 0.56 to 0.47 (16.1%) and from 0.57 to 0.45 (21.1%); NSE increased from 0.79 to 0.85 (7.6%) and from 0.80 to 0.87 (8.8%); and  $R^2$  increased from 0.87 to 0.91 (4.6%) and from 0.85 to 0.91 (7.1%), when using  $ET_{od}$  and  $ET_{osoh}$  as targets, respectively.

## 3.2 Local scenario

### 3.2.1 Temperature-based models

For the temperature-based models developed in local scenario, the results obtained for  $ET_{od}$  and  $ET_{soh}$  targets (Fig. 8) also generally had the same behavior. The best results were found for the models that used 24 h hourly data as input. Comparing the machine learning models, ANN performed a little better than RF and XGBoost. However, CNN had the best performance among all models, but with slight performance improvements. In relation to the ANN models developed with 24 h hourly data and hourly  $R_a$ , RMSE reduced from 0.44 to 0.41 (6.8%) and from 0.47 to 0.45 (4.3%); NSE increased from 0.88 to 0.89 (1.1%) and from 0.86 to 0.87 (1.2%); and  $R^2$  increased from 0.88 to 0.90 (2.3%) and from 0.87 to 0.89 (2.3%), when considering  $ET_{od}$  and  $ET_{soh}$  as targets, respectively.

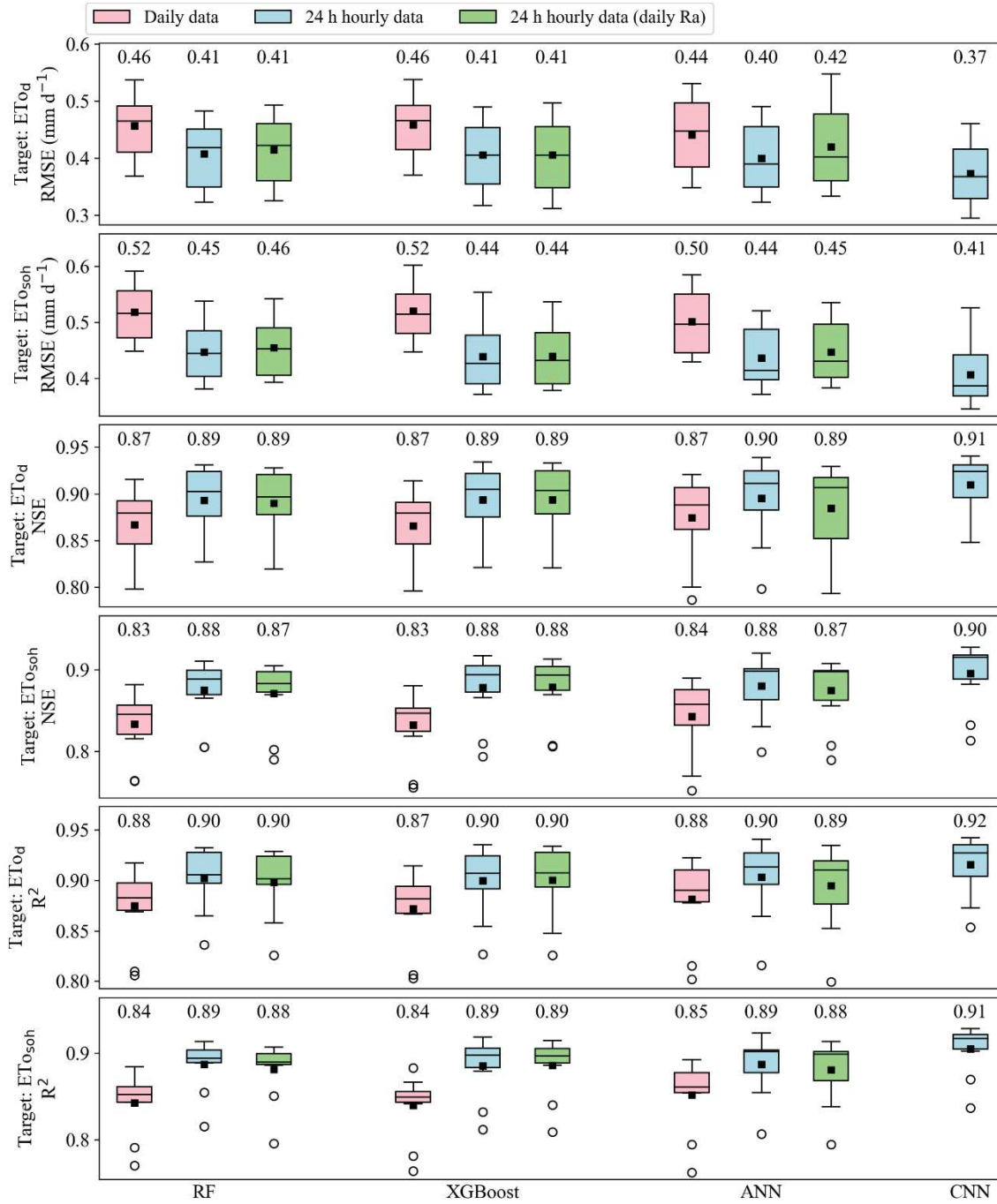


**Fig. 8.** Boxplots and mean values of RMSE, NSE and  $R^2$  for local temperature-based models developed using  $ET_{od}$  and  $ET_{soh}$  as target.

Comparing the best results, which were found using CNN (for both  $ET_{od}$  and  $ET_{soh}$  targets) developed using 24 h hourly data and hourly Ra, with the results found for the best model that used daily data (ANN, for both  $ET_{od}$  and  $ET_{soh}$  targets), RMSE reduced from 0.53 to 0.41 (22.6%) and from 0.58 to 0.45 (22.4%); NSE increased from 0.82 to 0.89 (8.5%) and from 0.79 to 0.87 (10.1%); and  $R^2$  increased from 0.83 to 0.90 (8.4%) and from 0.80 to 0.89 (11.3%), when using  $ET_{od}$  and  $ET_{soh}$  as targets, respectively.

### 3.2.2 Temperature- and relative humidity-based models

For the local temperature- and relative humidity-based models, for both  $ET_{od}$  and  $ET_{soh}$  targets the best results were found for the models that used 24 h hourly data as input with hourly or daily  $R_a$  (Fig. 9). Comparing the machine learning models, all the models showed similar performances, with a slight advantage for CNN. In relation to the ANN models developed with 24 h hourly data and hourly  $R_a$ , using CNN, RMSE reduced from 0.40 to 0.37 (7.5%) and from 0.44 to 0.41 (6.8%); NSE increased from 0.90 to 0.91 (1.1%) and from 0.88 to 0.90 (2.3%); and  $R^2$  increased from 0.90 to 0.92 (2.2%) and from 0.89 to 0.91 (2.2%), when considering  $ET_{od}$  and  $ET_{soh}$  as targets, respectively.



**Fig. 9.** Boxplots and mean values of RMSE, NSE and  $R^2$  for local temperature- and relative humidity-based models developed using  $ET_{od}$  and  $ET_{soh}$  as target.

Comparing the best results, which were found using CNN (for both  $ET_{od}$  and  $ET_{soh}$  targets) developed using 24 h hourly data and hourly Ra, with the results found for the best models that used daily data (ANN, for both  $ET_{od}$  and  $ET_{soh}$  targets), RMSE reduced from 0.44 to 0.37 (15.9%) and from 0.50 to 0.41 (18.0%); NSE increased from 0.87 to 0.91 (4.6%) and from 0.84 to 0.90 (7.1%); and  $R^2$  increased from 0.88 to 0.92 (4.5%) and from 0.85 to 0.91 (7.1%), when using  $ET_{od}$  and  $ET_{soh}$  as targets, respectively.



### 3.2.3 Overall evaluation

The use of hourly data measured during a 24 h period for estimation of daily ETo provided good results in all cases evaluated, outperforming models developed using daily input data (conventional approach). The use of hourly Ra in relation to daily Ra in the models that used 24 h hourly data did not promote large changes in the performance of the models. However, in almost all cases, the use of hourly Ra provided results equal or slightly higher than those obtained using daily Ra.

The best machine learning models obtained when daily data and hourly data with hourly Ra were considered as input, as well as their performance metrics, are presented in Table 1. It should be highlighted that when using temperature-based models developed with 24 h hourly data with hourly Ra; RMSE, NSE and  $R^2$  values were better than those obtained with the temperature- and relative humidity-based models developed with daily data.

**Table 1.** Best machine learning models and their respective RMSE, NSE and  $R^2$  mean values obtained when using daily data (Daily) and 24 h hourly data with hourly extraterrestrial radiation (Hourly) as input

Scenario	Data availability	Target	Input	Best model	RMSE	NSE	$R^2$
Regional	T	ET <sub>0d</sub>	Daily	ANN	0.70	0.68	0.82
		ET <sub>0d</sub>	Hourly	CNN	0.52	0.83	0.87
		ET <sub>0soh</sub>	Daily	ANN	0.71	0.69	0.79
		ET <sub>0soh</sub>	Hourly	CNN	0.51	0.84	0.88
	T+RH	ET <sub>0d</sub>	Daily	XGBoost	0.56	0.79	0.87
		ET <sub>0d</sub>	Hourly	CNN	0.47	0.85	0.91
		ET <sub>0soh</sub>	Daily	ANN	0.57	0.80	0.85
		ET <sub>0soh</sub>	Hourly	CNN	0.45	0.87	0.91
Local	T	ET <sub>0d</sub>	Daily	ANN	0.53	0.82	0.83
		ET <sub>0d</sub>	Hourly	CNN	0.41	0.89	0.90
		ET <sub>0soh</sub>	Daily	ANN	0.58	0.79	0.80
		ET <sub>0soh</sub>	Hourly	CNN	0.45	0.87	0.89
	T+RH	ET <sub>0d</sub>	Daily	ANN	0.44	0.87	0.88
		ET <sub>0d</sub>	Hourly	CNN	0.37	0.91	0.92

ET <sub>Osoh</sub>	Daily	ANN	0.50	0.84	0.85
ET <sub>Osoh</sub>	Hourly	CNN	0.41	0.90	0.91

T - temperature; RH - relative humidity.

In some cases, lower errors have been found when ET<sub>Od</sub> was used as target instead of ET<sub>Osoh</sub>, with RMSE differences ranging from 0.01 to 0.06 (Table 1). However, it is important to highlight that when comparing ET<sub>Od</sub> against ET<sub>Osoh</sub> (reference), RMSE values ranging from 0.24 to 0.32 were obtained (Table 2). Irmak et al. (2005) reported RMSE ranging from 0.25 to 0.56 (United States of America); Althoff et al. (2019) found mean RMSE equal to 0.22 (Brazil, including the study area of this study); and Djaman et al. (2018) found RMSE ranging from 0.36 to 0.84 (Africa). Therefore, assuming that ET<sub>Osoh</sub> is more accurate than ET<sub>Od</sub> (Allen et al., 1998; Althoff et al., 2019; Djaman et al., 2018; Irmak et al., 2005; Pereira et al., 2015) and that the RMSE differences obtained in this study when ET<sub>Od</sub> and ET<sub>Osoh</sub> were used as targets are much smaller than the ET<sub>Od</sub> errors showed above, it is recommended to use the machine learning models developed with ET<sub>Osoh</sub> as target.

**Table 2.** Comparison between ET<sub>Od</sub> and ET<sub>Osoh</sub> (reference) for test weather stations (whole dataset).

	ET <sub>Osoh</sub>	ET <sub>Od</sub>	RMSE	NSE	R <sup>2</sup>	MBE
Mean	3.83	3.93	0.28	0.95	0.96	0.10
Median	4.01	4.10	0.28	0.95	0.96	0.11
Min.	2.94	2.98	0.24	0.94	0.95	-0.06
Max.	4.47	4.54	0.32	0.96	0.98	0.21

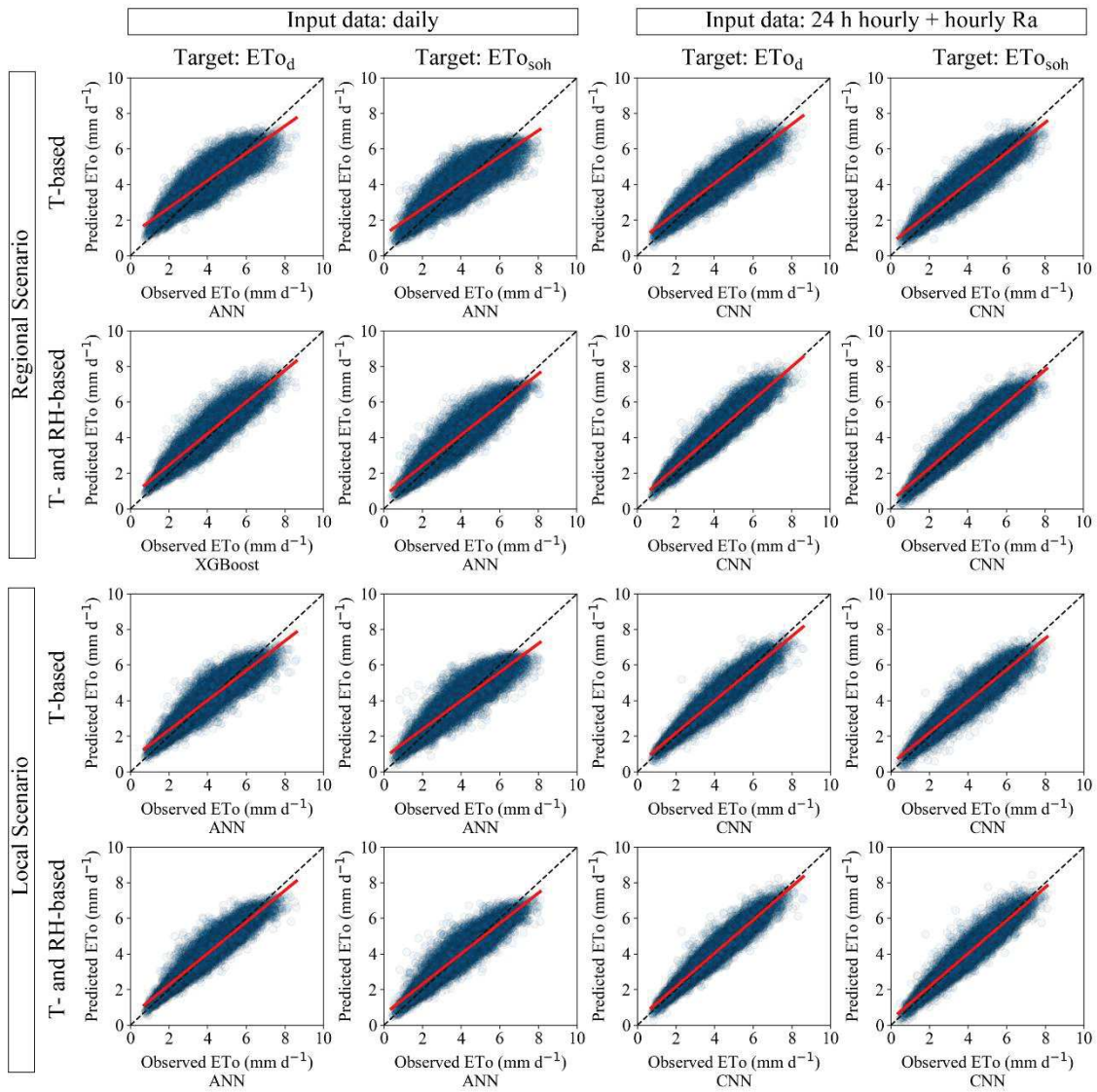
ET<sub>Osoh</sub>, ET<sub>Od</sub>, RMSE and MBE are expressed in mm d<sup>-1</sup>.

When ET<sub>Osoh</sub> was used as target, the models developed with 24 h hourly data provided performance gains over the models that used daily data (Table 1). This behavior is, to a certain extent, already expected since to obtain ET<sub>Osoh</sub> hourly data are needed. However, as the use of 24 h hourly data also provided gains when ET<sub>Od</sub> was used as target (Table 1), it can be concluded that this approach has a real potential to improve the estimation of ET<sub>O</sub> under limited data availability.

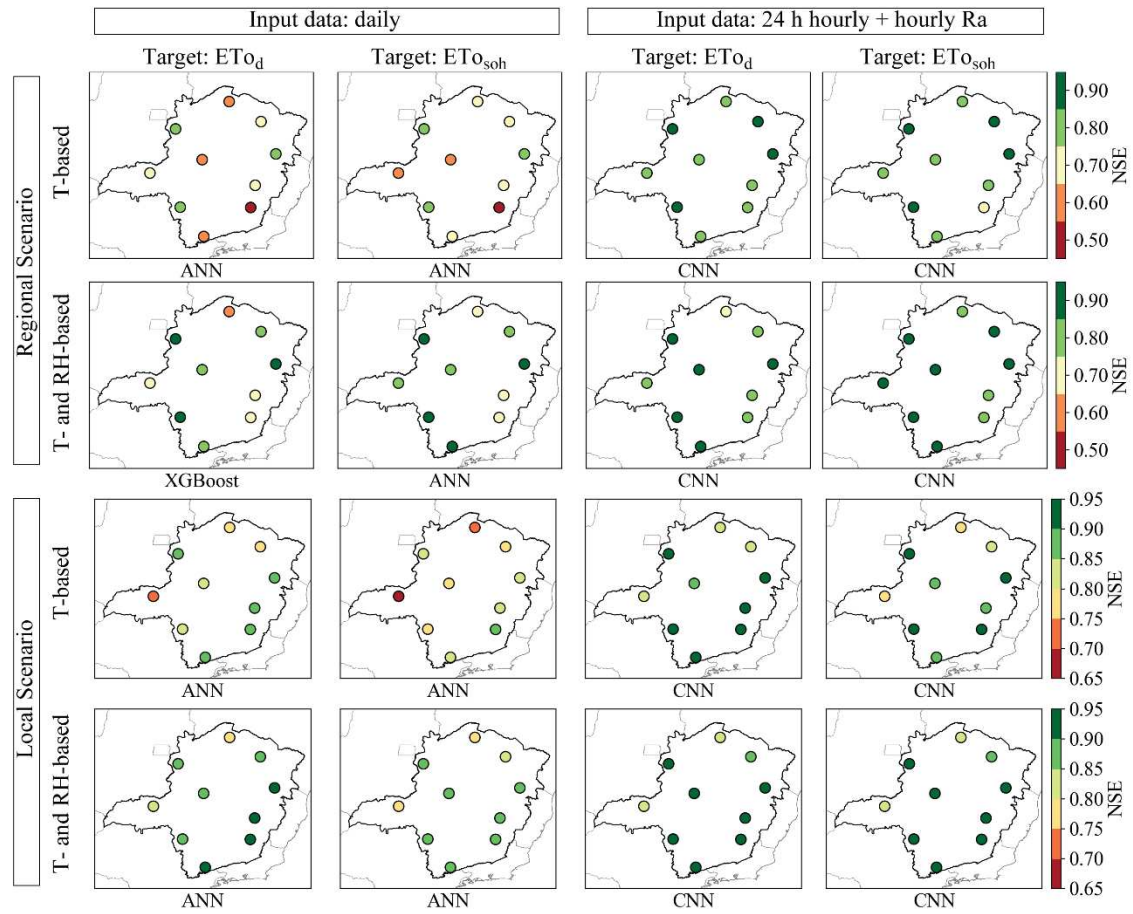
For the same input data, the models developed in local scenario had better performance than the models developed in regional scenario (Table 1). This behavior is very common since ET<sub>O</sub> modeling is more complex in regional scenario (Kiafar et al., 2017; Reis et al., 2019; Shiri

et al., 2014). However, it is worth mentioning that local models are specific for the climatic conditions of the place where they were developed. Therefore, regional models have a key role in places without full datasets to use the Penman-Monteith equation or to develop local models. Although local models typically perform better, in this study, the regional models developed using hourly data outperformed the local models developed with daily data, except in the case of the temperature- and relative humidity-based models with  $ET_{0d}$  target (Table 1).

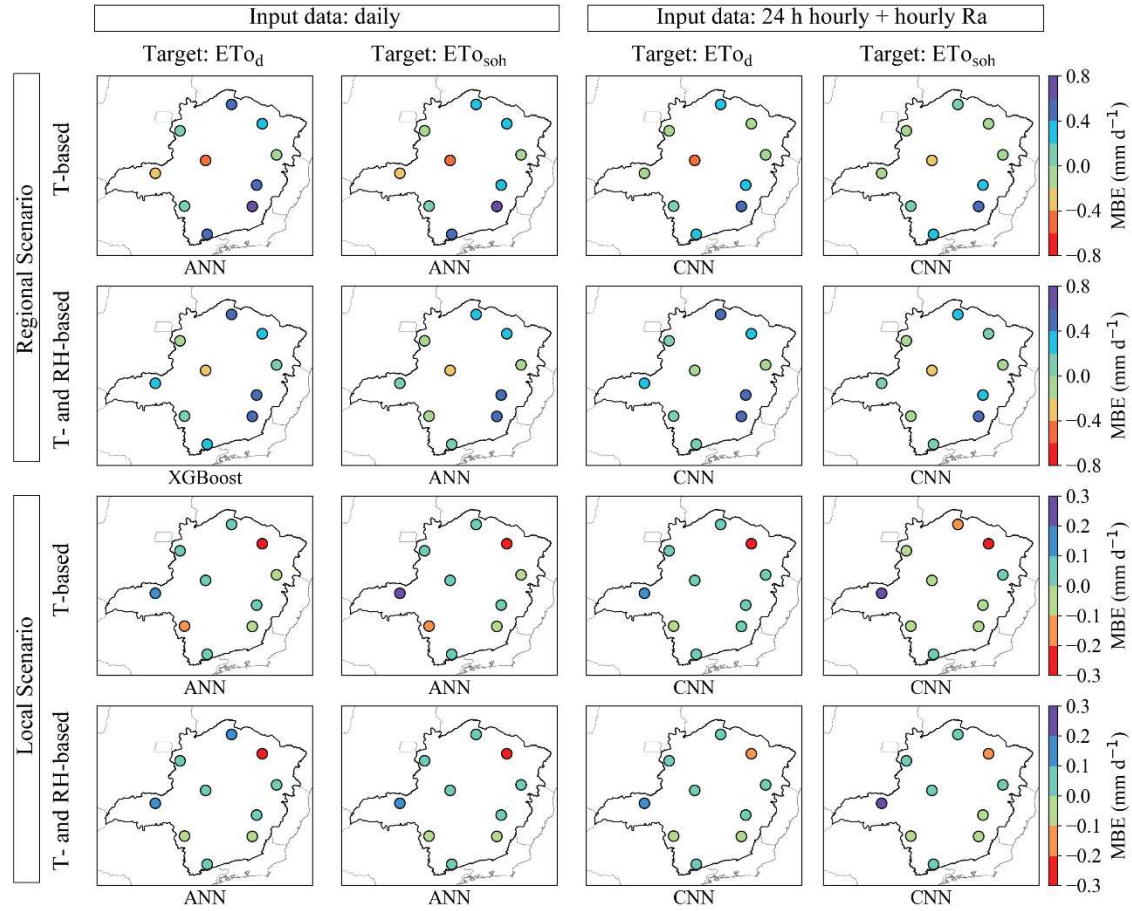
In addition to boxplots and mean values of performance metrics, it is also convenient to analyze scatter plots of predicted vs observed values and the behavior of the models in space. Thus, scatter plots and the distribution of NSE and MBE values in space for the machine learning models listed in Table 1 are presented for each test weather station in Figs. 10, 11 and 12, respectively. It can be noted that, when using 24 h hourly data, more accurate predictions were found (Fig. 10). In addition, in general, the use of 24 h hourly data provided performance gains, showing higher NSE values and lower MBE absolute values, in all the study region (Figs. 11 and 12). As the study area includes a wide range of climatic regimes (Figs. 2 and 3), it is indicated that the proposed approach performs well under different climatic conditions.



**Fig. 10.** Scatter plots for all test weather stations (pooled data) for the best regional and local machine learning models developed using daily data (daily) and 24 h hourly data with hourly extraterrestrial radiation (24 h hourly + hourly Ra) as input.



**Fig. 11.** Best regional and local machine learning models and their respective NSE values obtained for each weather station when using daily data (daily) and 24 h hourly data with hourly extraterrestrial radiation (24 h hourly + hourly Ra) as input.



**Fig. 12.** Best regional and local machine learning models and their respective MBE values obtained for each weather station when using daily data (daily) and 24 h hourly data with hourly extraterrestrial radiation (24 h hourly + hourly Ra) as input.

It is also important to note that the use of relative humidity generally promoted performance increments in all the study area, mainly for the models developed with daily input data. Although the role of relative humidity can be more important in arid or semi-arid regions, in the present study, this variable also provided performance gains in humid regions (southern Minas Gerais). Ferreira et al. (2019) also reported performance improvements when adding relative humidity to temperature in machine learning models developed for Brazil.

The performance gains obtained with the use of 24 h hourly data is probably because the models have captured patterns in the data measured along the day that can be associated with other variables not used as input. For instance, it is known that the difference between daily maximum and minimum air temperatures can be used as a cloud cover indicator, which is related to solar radiation (Allen et al., 1998; Hargreaves and Allen, 2003). Therefore, this and other patterns useful in ETo modeling can be empirically explored for machine learning models



when using 24 h hourly data, improving ETo modeling. This approach can also be evaluated in future studies for estimation of other meteorological variables.

When comparing the traditional machine learning models (i.e., RF, XGBoost and ANN), in general, they performed similar, but ANN models exhibited slightly higher performances. However, CNN outperformed all the models in all the cases studied, offering the best performance to estimate daily ETo. The highest performance gains promoted by the CNN models were found for the temperature-based models developed in regional scenario, in which, in relation to ANN, RMSE decreased by 11.9 and 8.9%; and NSE increased by 7.8 and 5.0%, for ETo<sub>d</sub> and ETo<sub>soh</sub> targets, respectively.

The good performance obtained for CNN is probably because it was able to extract features from the hourly input data that were not captured by the other models evaluated. The automatic feature extraction is a powerful characteristic of CNN, which allow it to identify the most useful features, which are discriminative information, from raw data, without the need for preprocessing (LeCun et al., 2015; Ng et al., 2019). The findings of this study corroborate other studies in which CNN has outperformed traditional machine learning models, reaching the best performance (Haidar and Verma, 2018; LeCun et al., 2015; Ng et al., 2019; Shen, 2018).

The use CNN to estimate daily ETo using 24 h hourly data and hourly Ra promoted great performance gains in relation to the use of traditional machine models with daily input data (conventional approach) (Table 1). Therefore, the proposed approach can be used to estimate ETo at a low cost and with good accuracy. In future studies, other CNN architectures can be explored. In addition, other deep learning models, such as Long short-term memory (LSTM) and Gated Recurrent Unit (GRU), can be assessed.

#### 4 Conclusions

This study assessed, for the first time, in regional and local scenarios, the use of limited hourly meteorological data (temperature and relative humidity or only temperature) to estimate daily ETo directly, employing RF, XGBoost, ANN and CNN. The following options were evaluated: (i) use of daily input data (conventional approach); (ii) use of hourly data measured during a 24 h period and hourly Ra to estimate daily ETo directly; and (iii) use of hourly data measured during a 24 h period and daily Ra to estimate daily ETo directly. All the options used Ra as input. To develop and evaluate the models, two types of targets were considered: ETo<sub>d</sub> (daily ETo computed using the daily version of ASCE-PM equation) and ETo<sub>soh</sub> (daily ETo computed by summing hourly ETo obtained with the ASCE-PM equation).

In all cases, the best results were found using hourly data to estimate daily ETo directly, with hourly or daily Ra. The use of limited 24 h hourly data to estimate daily ETo can enable

better estimates of ETo at a low cost. This idea can also be assessed in other places and to estimate other meteorological variables.

When compared to the models developed with daily data, in regional scenario, for temperature-based models, the use of 24 h hourly data + hourly Ra reduced RMSE by up to 28.2% and increased NSE and  $R^2$  by up to 21.7 and 11.4%, respectively. For temperature- and relative humidity-based models, RMSE reduced by up to 21.1% and NSE and  $R^2$  increased by up to 8.8 and 7.1%, respectively. In local scenario, for temperature-based models, RMSE reduced by up to 22.4% and NSE and  $R^2$  increased by up to 10.1 and 11.3%, respectively. For temperature- and relative humidity-based models, RMSE reduced by up to 18.0% and NSE and  $R^2$  increased by up to 7.1 and 7.1%, respectively.

In general, ANN performed slightly better than the other traditional machine learning models (i.e., RF and XGBoost), being the best model to estimate ETo with daily input data in almost all cases. However, CNN exhibited the best performance when hourly data were used as input, offering the most accurate ETo estimates. CNN models provided performance gains, in terms of RMSE and NSE, up to 11.9 and 7.8%, respectively, in relation to ANN models with the same input data. Future studies can explore the potential of other CNN architectures and other deep learning modes, such as LSTM and GRU. The models developed in this study are available upon request.

## Acknowledgments

The authors wish to thank the National Council for Scientific and Technological Development (CNPq) for providing a scholarship to the first author and the Brazilian National Institute of Meteorology (INMET) for the meteorological data used.

## References

- Abdeljaber, O., Avci, O., Kiranyaz, M.S., Boashash, B., Sodano, H., Inman, D.J., 2018. 1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data. *Neurocomputing*. 275, 1308-1317. doi:10.1016/j.neucom.2017.09.069
- Abdullah, S.S., Malek, M.A., Abdullah, N.S., Kisi, O., Yap, K.S., 2015. Extreme Learning Machines: A new approach for prediction of reference evapotranspiration. *J. Hydrol.* 527, 184-195. doi:10.1016/j.jhydrol.2015.04.073
- Adam-Bourdarios, C., Cowan, G., Germain, C., Guyon, I., Kégl, B., Rousseau, D., 2015. The Higgs boson machine learning challenge. In *NIPS 2014 Workshop on High-energy Physics and Machine Learning* (pp. 19-55).



Ahooghalandari, M., Khiadani, M., Jahromi, M.E., 2016. Developing Equations for Estimating Reference Evapotranspiration in Australia. *Water Resour. Manag.* 30, 3815–3828. doi:10.1007/s11269-016-1386-7

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and Drainage, Paper No. 56, Food and Agriculture Organization of the United Nations, Rome.

Allen, R.G., Pruitt, W.O., Wright, J.L., Howell, T.A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J.B., Smith, M., Pereira, L.S., Raes, D., Perrier, A., Alves, I., Walter, I., Elliott, R., 2006. A recommendation on standardized surface resistance for hourly calculation of reference ET o by the FAO56 Penman-Monteith method. *Agric. Water Manag.* 81, 1-22. doi:10.1016/j.agwat.2005.03.007

Almorox, J., Senatore, A., Quej, V.H., Mendicino, G., 2018. Worldwide assessment of the Penman–Monteith temperature approach for the estimation of monthly reference evapotranspiration. *Theor. Appl. Climatol.* 131, 693–703. doi:10.1007/s00704-016-1996-2

Althoff, D., Filgueiras, R., Dias, S.H.B., Rodrigues, L.N., 2019. Impact of sum-of-hourly and daily timesteps in the computations of reference evapotranspiration across the Brazilian territory. *Agric. Water Manag.* 226, 105785. doi:10.1016/j.agwat.2019.105785

Alvares, C.A., Stape, J.L., Sentelhas, P.C., De Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen’s climate classification map for Brazil. *Meteorol. Zeitschrift.* 22, 711–728. doi:10.1127/0941-2948/2013/0507

ASCE-EWRI, 2005. The ASCE standardized reference evapotranspiration equation. In: Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E., Snyder, R.L. (Eds.), Report 0-7844-0805-X. Am. Soc. Civ. Eng.-Environ. Water Resources Instit., 69 pp (+App. A-F and Index).

Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5-32. doi:10.1023/A:1010933404324

Canizo, M., Triguero, I., Conde, A., Onieva, E., 2019. Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study. *Neurocomputing.* 363, 246-260. doi:10.1016/j.neucom.2019.07.034

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. doi:10.1145/2939672.2939785

Djaman, K., Irmak, S., Sall, M., Sow, A., & Kabenge, I. (2018). Comparison of sum-of-hourly and daily time step standardized ASCE Penman-Monteith reference evapotranspiration. *Theor. Appl. Climatol.* 134(1-2), 533-543. doi:10.1007/s00704-017-2291-6

Dyrmann, M., Karstoft, H., Midtiby, H.S., 2016. Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72-80. doi:10.1016/j.biosystemseng.2016.08.024

Exner-Kittridge, M., 2012. Closure to “case study on the accuracy and cost/effectiveness in simulating reference evapotranspiration in West-Central Florida” by Michael Grant Exner-Kittridge and Mark Cable Rains. *J. Hydrol. Eng.* 17, 225-226. doi:10.1061/(ASCE)HE.1943-5584.0000430

Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., Zeng, W., 2019. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric. Water Manag.* 225, 105758. doi:10.1016/j.agwat.2019.105758

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., Xiang, Y., 2018. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* 263, 225-241. doi:10.1016/j.agrformet.2018.08.019

Feng, Y., Cui, N., Gong, D., Zhang, Q., Zhao, L., 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agric. Water Manag.* 193, 163-173. doi:10.1016/j.agwat.2017.08.003

Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. *J. Hydrol.* 572, 556–570. doi:10.1016/j.jhydrol.2019.03.028

Gao, S., Wang, X., Miao, X., Su, C., Li, Y., 2019. ASM1D-GAN: An Intelligent Fault Diagnosis Method Based on Assembled 1D Convolutional Neural Network and Generative Adversarial Networks. *J. Signal Process. Syst.* 91, 1237-1247. doi:10.1007/s11265-019-01463-8

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press.

Haidar, A., Verma, B., 2018. Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network. *IEEE Access.* 6, 69053-69063. doi:10.1109/ACCESS.2018.2880044

Hargreaves, G.H., Allen, R.G., 2003. History and evaluation of Hargreaves evapotranspiration equation. *J. Irrig. Drain. Eng.* 129, 53–63. doi:10.1061/(ASCE)0733-9437(2003)129:1(53)

Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. *Appl. Eng. Agric.* 1 (2), 96–99.

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* 574, 1029-1041. doi:10.1016/j.jhydrol.2019.04.085

Irmak, S., Howell, T. A., Allen, R. G., Payero, J. O., & Martin, D. L., 2005. Standardized ASCE Penman-Monteith: Impact of sum-of-hourly vs. 24-hour timestep computations at reference weather station sites. *Trans. ASAE.* 48(3), 1063-1077. doi:10.13031/2013.18517Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* 10, 75. doi:10.3390/rs10010075

Kamilaris, A., Prenafeta-Boldú, F.X., 2018a. A review of the use of convolutional neural networks in agriculture. *J. Agric. Sci.* 153, 312-322. doi:10.1017/S0021859618000436

Kamilaris, A., Prenafeta-Boldú, F.X., 2018b. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70-90. doi:10.1016/j.compag.2018.02.016

Kiafar, H., Babazadeh, H., Marti, P., Kisi, O., Landaras, G., Karimi, S., Shiri, J., 2017. Evaluating the generalizability of GEP models for estimating reference evapotranspiration in distant humid and arid locations. *Theor. Appl. Climatol.* 130, 377-389. doi:10.1007/s00704-016-1888-5

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.*

Kumar, M., Raghuwanshi, N.S., Singh, R., 2011. Artificial neural networks approach in evapotranspiration modeling: A review. *Irrig. Sci.* 29, 11-25. doi:10.1007/s00271-010-0230-8

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature.* 521, 436. doi:10.1038/nature14539

Li, D., Zhang, J., Zhang, Q., Wei, X., 2017. Classification of ECG signals based on 1D convolution neural network, in: 2017 IEEE 19th International Conference on E-Health Networking, Applications and Services, Healthcom 2017. doi:10.1109/HealthCom.2017.8210784

Liu, L., Ji, M., Buchroithner, M., 2018. Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery. *Sensors (Switzerland)*. 18, 3169. doi:10.3390/s18093169

Mehdizadeh, S., Behmanesh, J., Khalili, K., 2017. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput. Electron. Agric.* 139, 103–114. doi:10.1016/j.compag.2017.05.002

Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*. 352, 251-267. doi: 10.1016/j.geoderma.2019.06.016

Nourani, V., Elkiran, G., Abdullahi, J., 2019. Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *J. Hydrol.* 577, 123958. doi:10.1016/j.jhydrol.2019.123958

Paredes, P., Pereira, L.S., 2019. Computing FAO56 reference grass evapotranspiration PM-ET o from temperature with focus on solar radiation. *Agric. Water Manag.* 215, 86-102. doi:10.1016/j.agwat.2018.12.014

Pereira, L.S., Allen, R.G., Smith, M., Raes, D., 2015. Crop evapotranspiration estimation with FAO56: Past and future. *Agric. Water Manag.* 147, 4-20. doi:10.1016/j.agwat.2014.07.031

Reis, M.M., da Silva, A.J., Zullo Junior, J., Tuffi Santos, L.D., Azevedo, A.M., Lopes, É.M.G., 2019. Empirical and learning machine approaches to estimating reference evapotranspiration based on temperature data. *Comput. Electron. Agric.* 165, 104937. doi:10.1016/j.compag.2019.104937

Shen, C., 2018. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resour. Res.* 54, 8558-8593. doi:10.1029/2018WR022643

Shiri, J., Nazemi, A.H., Sadraddini, A.A., Landaras, G., Kisi, O., Fakheri Fard, A., Marti, P., 2014. Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in Iran. *Comput. Electron. Agric.* 108, 230–241. doi:10.1016/j.compag.2014.08.007

Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water (Switzerland)*. 11, 910. doi:10.3390/w11050910

UNEP (United Nations Environmental Programme). 1997. World Atlas of Desertification. Editorial commentary by N. Middleton and D.S.G. Thomas, London, Edward Arnold.

Valiantzas, J.D., 2018. Temperature-and humidity-based simplified Penman's ET<sub>0</sub> formulae. Comparisons with temperature-based Hargreaves-Samani and other methodologies. *Agric. Water Manag.* 208, 326-334. doi:10.1016/j.agwat.2018.06.028

Valiantzas, J.D., 2012. Discussion of “case study on the accuracy and cost/effectiveness in simulating reference evapotranspiration in West-Central Florida” by Michael Grant Exner-Kittridge and mark cable rains. *J. Hydrol. Eng.* 17, 224–225. Doi:10.1061/(ASCE)HE.1943-5584.0000394

Wang, S., Lian, J., Peng, Y., Hu, B., Chen, H., 2019. Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China. *Agric. Water Manag.* 221, 220-230. doi:10.1016/j.agwat.2019.03.027

Wu, L., Fan, J., 2019. Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. *PLoS One.* 14, e0217520. doi:10.1371/journal.pone.0217520

Zanetti, S.S., Dohler, R.E., Cecílio, R.A., Pezzopane, J.E.M., Xavier, A.C., 2019. Proposal for the use of daily thermal amplitude for the calibration of the Hargreaves-Samani equation. *J. Hydrol.* 571, 193-201. doi:10.1016/j.jhydrol.2019.01.049

## Article 2: Multi-step ahead forecasting of daily reference evapotranspiration using deep learning

**Abstract:** Daily reference evapotranspiration (ET<sub>o</sub>) forecasts can help farmers in irrigation planning. Therefore, this study assesses the potential of deep learning (long short-term memory (LSTM), one-dimensional convolutional neural network (1D CNN) and a combination of the two previous models (CNN-LSTM)) and traditional machine learning models (artificial neural network (ANN) and random forest (RF)), in regional and local scenarios, to forecast multi-step ahead daily ET<sub>o</sub> (seven days) using iterated, direct and multiple input multiple output (MIMO) forecasting strategies. Three input data combinations were assessed: (1) only lagged ET<sub>o</sub>; (2) lagged ET<sub>o</sub> + day of the year of each step of the time lag considered; and (3) the same of input combination 2 + lagged meteorological variables. Data from 53 weather stations located in Minas Gerais, Brazil, were used. Four stations were used as test stations. Two baselines were also employed: (B1), all the forecasting horizon is considered equal to the mean ET<sub>o</sub> measured during the last seven days; and (B2), ahead ET<sub>o</sub> values are considered equal to their respective historical monthly means. In general, MIMO was the best forecasting strategy, offering good performance and lower computational cost. The deep learning models performed slightly better than the machine learning models, and both were better than the best baseline (B2), mainly on the first and second forecasting days. Among the deep learning models, CNN-LSTM2 (i.e., CNN-LSTM with input combination 2) performed the best in local scenario (mean RMSE over the prediction horizon and stations equal to 0.87), and CNN-LSTM3 performed the best in regional scenario (mean RMSE equal to 0.88). The regional models are recommended instead of the local models since they exhibited similar performances and have higher generalization capacity. Finally, although the models developed have not exhibited high accuracies, they can be useful tools in places where historical monthly mean ET<sub>o</sub> is used to forecast ET<sub>o</sub>.

**Keywords:** long short-term memory, one-dimensional convolutional neural network, machine learning, time series

### 1 Introduction

Computation of evapotranspiration (ET) is required in water resource management, hydrological studies, irrigation scheduling and crop modeling studies. In addition, some drought indices are computed based on ET (Tian et al., 2020). There are different alternatives to estimate ET of a particular crop, however, the use of reference evapotranspiration (ET<sub>o</sub>) and crop coefficient (K<sub>c</sub>) is the most common approach (Pereira et al., 2015). ET<sub>o</sub> can be estimated

using the FAO-56 Penman-Monteith (FAO56-PM) equation, which is widely used and accepted due to its good performance (Allen et al., 1998; Pereira et al., 2015).

In addition to estimation of ETo from past and current periods, prediction of ahead ETo values (i.e., ETo forecasting) can be useful for irrigation planning. This future information can improve real-time irrigation scheduling, allowing to make better decisions. However, ETo forecasting studies are not as common in the literature as ETo estimation studies.

Trajkovic et al. (2003) forecasted monthly ETo based on previous ETo values using artificial neural network (ANN). Landeras et al. (2009) employed autoregressive integrated moving average (ARIMA) and ANN to forecast weekly ETo. Torres et al. (2011) forecasted daily ETo (computed using the Hargreaves-Samani equation) up to seven days ahead using ANN and multivariate relevance vector machine (MVRVM). Karbasi (2018) forecasted multi-step ahead daily ETo using gaussian process regression (GPR) and wavelet-GPR. The author observed that by increasing the forecasting time period from 1 to 30 days, the accuracy was reduced. Ashrafzadeh et al. (2020) forecasted monthly ETo up to 24 months ahead using seasonal autoregressive integrated moving average (SARIMA), group method of data handling (GMDH) and SVM. Nourani et al. (2020) forecasted multi-step ahead ETo based on monthly lagged meteorological data (up to 12 months). The authors used support vector machine (SVM), adaptive neuro fuzzy inference system (ANFIS), ANN and multiple linear regression (MLR) individually and in a multi-model approach (ensemble). The multi-model approach provided the best performance. In Brazil, Alves et al. (2017) forecasted one-day ahead daily ETo with high accuracy using ANN with only mean air temperature from the previous day as input.

In contrast to the studies presented above, some studies forecast ETo based on forecasted meteorological data, such as public weather forecasts (Cai et al., 2007; Perera et al., 2014; Traore et al., 2017; Yang et al., 2019), or with a combination of past local meteorological data and forecasted meteorological data (Bachour et al., 2016). In this study, ETo is forecasted based on past meteorological data. This approach has the advantage of not requiring external data, using only data measured on a local weather station.

The most common approach to forecast ETo with machine learning is to use only lagged ETo values as input data (Ashrafzadeh et al., 2020; Landeras et al., 2009; Trajkovic et al., 2003). However, the use of meteorological data related to ETo (i.e., temperature, relative humidity, solar radiation, wind speed and extraterrestrial radiation) and the day of the year as additional input data could provide performance gains. Thus, this is also investigated in the present study.

As presented above, traditional machine learning models, such as ANN and SVM, have been used for ETo forecasting. However, deep learning models can also be used for this task.



The deep learning field has gained much attention in recent years and has been applied in several areas, outperforming traditional machine learning models and achieving state-of-the-art performances (Ferreira and Cunha, 2020; Gao et al., 2019; Haidar and Verma, 2018; Kamilaris and Prenafeta-Boldú, 2018; Lecun et al., 2015; Lee et al., 2020; Saggi and Jain, 2019). A review of deep learning for water resources scientists is presented by Shen (2018). For time series forecasting, long short-term memory (LSTM) (Son and Kim, 2020; Tian et al., 2018; Zhou et al., 2019) and one-dimensional convolutional neural network (1D CNN) (Amarasinghe et al., 2017; Barzegar et al., 2020; Sayeed et al., 2020; Tian et al., 2018) can be used. In addition, 1D CNN can be combined with LSTM, creating a hybrid model (CNN-LSTM) (Barzegar et al., 2020; Kim and Cho, 2019; Huang and Kuo, 2018). Although deep learning has exhibited great performance in several cases, it is still poorly explored in the hydrology/climatology field. For ETo forecasting, studies using deep learning are very scarce.

For multi-step ahead forecasting, different modeling strategies can be employed (Taieb et al., 2010; Taieb et al., 2012; Ye and Dai, 2019). Iterated and direct strategies are the most common alternatives (Ye and Dai, 2019). In iterated strategy, a model is built to perform a one-step ahead forecasting. The result is fed back as input to predict the following value until the desired prediction horizon has been reached. In direct strategy, to predict the  $H$  next values of the time series,  $H$  models are constructed. Each model predicts a specific value of the prediction horizon. In contrast to iterated strategy, this method avoids accumulation of prediction errors. However, it has higher computational cost since more models are constructed.

In addition to iterated and direct strategies, multiple input multiple output (MIMO) strategy can be used (Taieb et al., 2010; Taieb et al., 2012; Ye and Dai, 2019). In MIMO, all the  $H$  next values of the time series are predicted at the same time. Thus, instead of a scalar value, a vector of future values is predicted. The advantages of MIMO against iterated and direct strategies are that it avoids accumulation of prediction errors and requires only a single model. In addition, compared to direct strategy, MIMO preserves the stochastic dependency characterizing the forecasted time series. However, to use MIMO, models that can predict multiple values, such as ANN and random forest (RF), are required. According to our knowledge, so far, multiple forecasting strategies have not been compared in ETo forecasting studies.

Considering the importance of ETo forecasts, this study assesses the potential of deep learning models (LSTM, 1D CNN and CNN-LSTM) and traditional machine learning models (ANN and RF), in regional and local scenarios, to forecast multi-step ahead daily ETo (seven



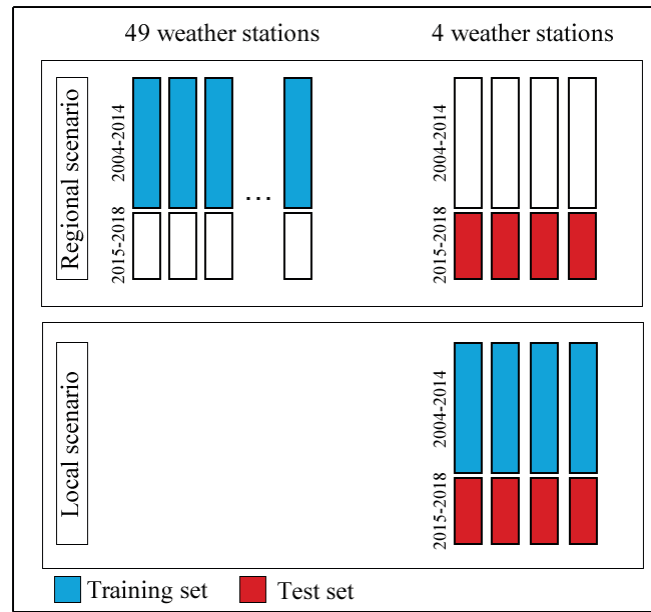
days) using iterated, direct and MIMO forecasting strategies. The use of additional input variables to lagged ETo (meteorological data and day of the year) is also assessed.

## **2 Materials and methods**

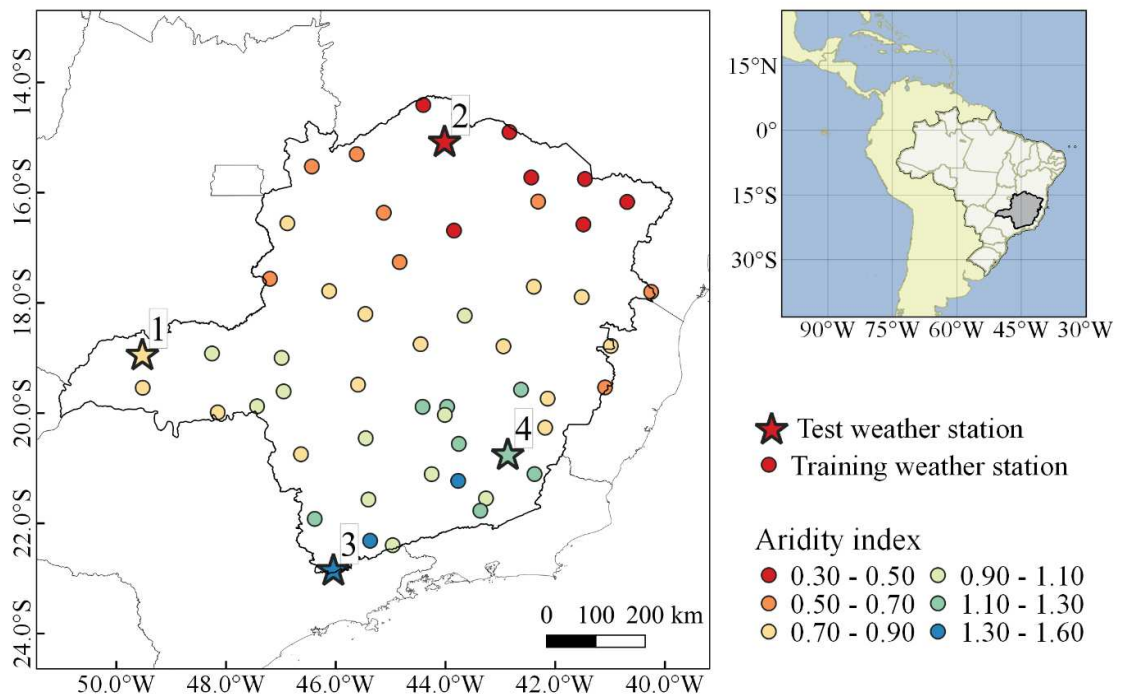
### **2.1 Database, study sites and data management**

Data from 53 automatic weather stations of the Brazilian National Institute of Meteorology (INMET) were used. The data length varied according to the stations, with a mean length of  $11.7 \pm 1.34$  years. All the stations have data up to the year 2018. Maximum and minimum air temperature, maximum and minimum relative humidity, solar radiation and wind speed were obtained. Wind speed, measured at 10 m height, was converted to 2 m, according to Allen et al. (1998). The data, collected on an hourly timescale, were converted to a daily timescale. Days with missing or faulty data were removed. Faulty data were identified on a daily timescale when minimum temperature was higher than maximum temperature; relative humidity out of the range 0-100%; minimum relative humidity higher than maximum relative humidity; negative wind speed; negative solar radiation; or solar radiation higher than extraterrestrial radiation.

To carry out the study, two scenarios were assessed: local and regional (Fig. 1). In local scenario, models were trained and evaluated individually in each weather station. In this scenario, four stations (test stations) were used. In regional scenario, models were trained using pooled data from 49 weather stations (training stations) and evaluated individually in four independent weather stations (the same stations evaluated in local scenario). Only four test stations were used in both scenarios to allow a detailed presentation of the results obtained for each station. To be more representative, stations with different climate conditions were selected. In both scenarios, a time split was also considered, using data up to 2014 to train the models and data from 2015 to 2018 to test them. The location and aridity index (IA) (UNEP (United Nations Environmental Programme), 1997) (ratio between annual precipitation to annual reference evapotranspiration), as well as the training/test split of the weather stations, are presented in Fig. 2. Detailed information on the stations used is presented in Table 1.



**Fig. 1.** Data management scenarios and training/test split used.



**Fig. 2.** Location, aridity index and training/test split of the weather stations used in the study. Test stations are followed by their identification number.

**Table 1.** Mean and standard deviation of daily climate characteristics and data length for the weather stations used.

Training period					
	Ituiutaba (1)	Mocambinho (2)	Monte Verde (3)	Viçosa (4)	Other 49 stations
Period	2006-2014	2007-2014	2004-2014	2005-2014	-
Samples	2892	2369	2946	2773	125453
T <sub>max</sub>	31.2 (±3)	32.0 (±2.6)	21.4 (±3.3)	26.2 (±3.4)	28.2 (±3.8)
T <sub>min</sub>	17.7 (±3.6)	18.7 (±3)	10.0 (±4.2)	14.8 (±3.6)	16.7 (±3.7)
RH <sub>max</sub>	91.0 (±9.2)	88.2 (±8.2)	95.4 (±1.9)	95.5 (±3)	89.6 (±9.4)
RH <sub>min</sub>	38.1 (±15.1)	35.5 (±14.1)	54.6 (±15.2)	50.1 (±13.5)	43.4 (±15.4)
WS	1.2 (±0.6)	0.9 (±0.3)	1.0 (±0.6)	0.7 (±0.3)	1.5 (±0.7)
SR	18.9 (±4.9)	21.1 (±4.9)	16.7 (±6.1)	15.6 (±5.7)	18.7 (±5.8)
ET <sub>o</sub>	4.2 (±1.2)	4.4 (±1.1)	2.9 (±1.1)	3.0 (±1.1)	4.1 (±1.4)
Test period					
	Ituiutaba (1)	Mocambinho (2)	Monte Verde (3)	Viçosa (4)	Other 49 stations
Period	2015-2018	2015-2018	2015-2018	2015-2018	-
Samples	1297	1354	1341	1456	62642
T <sub>max</sub>	31.5 (±3.1)	33.2 (±2.8)	21.9 (±3.4)	27.4 (±3.5)	28.9 (±4)
T <sub>min</sub>	17.9 (±3.6)	18.9 (±2.8)	10.2 (±4.2)	15.7 (±3.3)	17.1 (±3.6)
RH <sub>max</sub>	88.2 (±7.9)	85.3 (±7.7)	97.4 (±2.2)	95.4 (±2.5)	88.2 (±10.1)
RH <sub>min</sub>	39.7 (±15.3)	30.4 (±12.6)	54.1 (±16.3)	49.8 (±13.4)	41.7 (±15.1)
WS	1.1 (±0.5)	0.8 (±0.3)	1.1 (±0.4)	0.6 (±0.3)	1.4 (±0.7)
SR	18.9 (±4.9)	21.7 (±5.1)	16.7 (±6.2)	16.0 (±5.9)	18.7 (±5.9)
ET <sub>o</sub>	4.2 (±1.2)	4.6 (±1.1)	3.0 (±1.1)	3.2 (±1.2)	4.1 (±1.4)

T<sub>max</sub> - maximum air temperature, °C; T<sub>min</sub> - minimum air temperature, °C; RH<sub>max</sub> - maximum relative humidity, %; T<sub>max</sub> - maximum relative humidity, %; WS - wind speed, m s<sup>-1</sup>; SR - solar radiation, MJ m<sup>-2</sup> d<sup>-1</sup>; ET<sub>o</sub> - daily reference evapotranspiration, mm d<sup>-1</sup>.

## 2.2 Computation of reference evapotranspiration

To compute ET<sub>o</sub>, which was used to develop and evaluate the models under study, the FAO56-PM equation was used (Eq. 1). All procedures were performed according to Allen et al. (1998). The FAO56-PM equation was employed in the present study because it is widely used and accepted due to its good performance (Allen et al., 1998; Pereira et al., 2015).

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

where  $ET_o$  is the reference evapotranspiration ( $\text{mm d}^{-1}$ ),  $R_n$  is the net solar radiation ( $\text{MJ m}^{-2} \text{d}^{-1}$ ),  $G$  is the soil heat flux ( $\text{MJ m}^{-2} \text{d}^{-1}$ ) (considered to be null for daily estimates),  $T_{avg}$  is the daily mean air temperature ( $^{\circ}\text{C}$ ),  $u_2$  is the wind speed at a 2 m height ( $\text{m s}^{-1}$ ),  $e_s$  is the saturation vapor pressure (kPa),  $e_a$  is the actual vapor pressure (kPa) (obtained using maximum and minimum relative humidity),  $\Delta$  is the slope of the saturation vapor pressure function ( $\text{kPa } ^{\circ}\text{C}^{-1}$ ), and  $\gamma$  is the psychrometric constant ( $\text{kPa } ^{\circ}\text{C}^{-1}$ ).

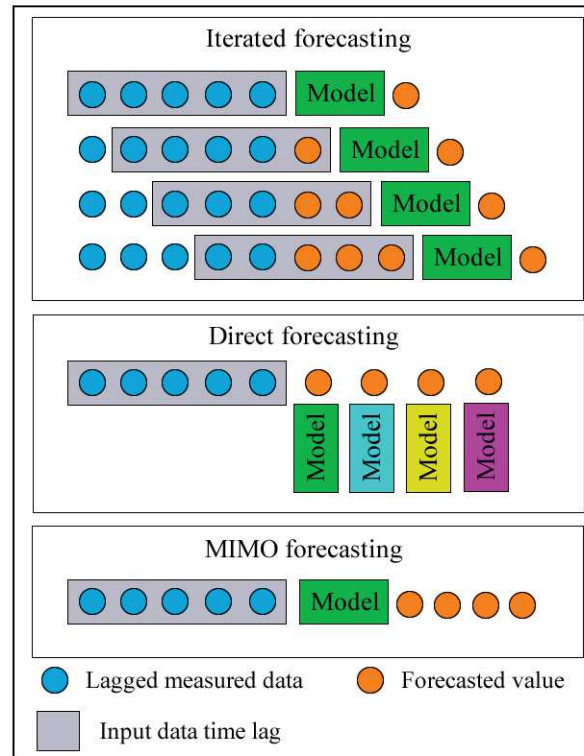
### 2.3 Forecasting scheme and input data combinations

To train the models and perform the optimization of input data time lag and hyperparameters, the data available to train the models were split into two subsets: training (data up to December 1, 2013) and validation (December 2, 2013 - December 31, 2014). Finally, the performance of the models was evaluated using data from the test set (2015-2018). In this study, the use of data from up to 30 lagged days was assessed (described in detail below). Thus, to ensure that the models developed with data from different numbers of lagged days are trained, validated and tested with the same dataset, the first 30 days from all the datasets (i.e., training, validation and test) were reserved to be used as lagged input data. For instance, all the models were validated for the period from January 1, 2014 to December 31, 2014, with the data from December 2, 2013 to December 31, 2013 reserved to be used as input data.

To forecast  $ET_o$ , seven days ahead were considered as prediction horizon. This period was considered because it is adequate for irrigation planning. To develop the machine learning and deep learning models, three input data combinations were assessed: (1) only lagged  $ET_o$ ; (2) lagged  $ET_o$  + day of the year of each step of the time lag considered; and (3) lagged  $ET_o$  + day of the year of each step of the time lag considered + lagged meteorological variables used to compute  $ET_o$  (maximum and minimum temperature, maximum and minimum relative humidity, wind speed, solar radiation and extraterrestrial radiation). The models developed using input data combinations 1, 2 and 3 were identified with the input combination number after their names (e.g., ANN1, ANN2 and ANN3).

Input data time lag was selected by trial and error, evaluating the use of data from up to 30 past days. In the case of traditional machine learning models, 1, 2, 3, 4, 5, 10, 15, 20, 25 and 30 past days were tested. In the case of deep learning models, 5, 10, 15, 20, 25 and 30 past days were tested. The optimal time lag was defined as the lag that promoted the lowest prediction error (lowest root mean square error) in the validation subset.

Three forecasting strategies were used to develop the models: iterated, direct and multiple input multiple output (MIMO) (Fig. 3). In iterated strategy, the first step ahead is forecasted and the result is fed back as input to predict the following value. This process is repeated until the desired prediction horizon has been reached. In direct strategy, H models are built to predict the H values of the prediction horizon. In MIMO, all the H values of the prediction horizon are predicted at the same time using a single model.



**Fig. 3.** Forecasting strategies used in the study.

## 2.4 Baselines

When developing models to forecast ETo, it can be useful to compare their performances against some baseline approaches. Therefore, two baselines were defined in this study (Table 2). The first baseline (B1) considers all the values of the prediction horizon ( $ET_{t+1}$ ,  $ET_{t+2}$ , ...,  $ET_{t+7}$ ) equal to the mean ETo obtained during the last seven days ( $ET_{t-6}$ ,  $ET_{t-5}$ , ...,  $ET_t$ ); and the second baseline (B2) forecasts the prediction horizon considering ahead ETo values equal to their respective historical monthly means (computed using the training dataset).

**Table 2.** Baselines used to forecast ETo.

Baseline	Prediction horizon / Forecasting rule			
	$ET_{t+1}$	$ET_{t+2}$	...	$ET_{t+7}$
B1	$ET_{m7}$	$ET_{m7}$	...	$ET_{m7}$
B2	$ET_{mm}$	$ET_{mm}$	...	$ET_{mm}$

$ET_{m7}$  - mean ETo measured during the last seven days ( $ET_{t-6}$ ,  $ET_{t-5}$ , ...,  $ET_t$ );  $ET_{mm}$  - respective mean monthly ETo for the steps of the prediction horizon.

## 2.5 Traditional machine learning and deep learning models

Two traditional machine learning models, RF and ANN, as well as three deep learning models, LSTM, 1D CNN and CNN-LSTM, were used. To implement the models, the Python programming language was used, employing the following libraries: Scikit-learn 0.22.1 and TensorFlow 2.1.0. A virtual machine from Google Cloud Platform was used for computations. For deep learning models training, a virtual machine with a graphics processing unit (GPU) was used.

The models were initially trained for all the data input combinations, input data lags and forecasting strategies mentioned above without hyperparameter tuning. Finally, the best performing models (i.e., models with optimal input data lag and optimal forecasting strategy) were submitted to hyperparameter tuning, selecting hyperparameters values that promoted the lowest prediction errors in the validation subset. Hyperparameter tuning was performed by grid search for machine learning models and by random search for deep learning models. Random search with 35 iterations was used to optimize deep learning models due to the higher number of hyperparameters to be optimized, which makes grid search highly computationally expensive.

### 2.5.1 Traditional machine learning models

#### 2.5.1.1 Random forest (RF)

RF, proposed by Breiman (2001), is a decision tree-based algorithm. In this algorithm, decision trees are fitted in different subsets of the training data. Although each individual tree is considered a weak learner, RF combines the prediction of all trees (ensemble), creating a powerful model (Huang et al., 2019). An advantage of this model is that it generally requires less adjustment in hyperparameters. More details about RF can be found in Silva Júnior et al. (2019) and Tyrallis et al. (2019).

To identify the optimal input data lag and optimal forecasting strategy, a RF was trained using 200 trees and the default values of the other hyperparameters. During hyperparameter

tuning phase, done for the best performing models (i.e., those with optimal input data lag and optimal forecasting strategy), the following hyperparameters and their respective values were assessed: `n_estimators` (number of trees) (100, 200, 400 and 600); `max_features` (number of features considered for splitting at each leaf node) (all features, the square root of the total number of features, and one third of the features) and `min_samples_leaf` (minimum number of samples required to be at a leaf node) (1, 5, 10 and 15).

#### **2.5.1.2 Artificial neural network (ANN)**

ANN of the feed-forward multilayer perceptron (MLP) type was used. ANN is capable of capturing complex relations between a set of input variables and one or more output variables. ANN is composed of artificial neurons organized in layers, which are connected to input data and other neurons. All the connections are made using weights, which are defined during the training process. More information on ANN can be found in Ferreira et al. (2019).

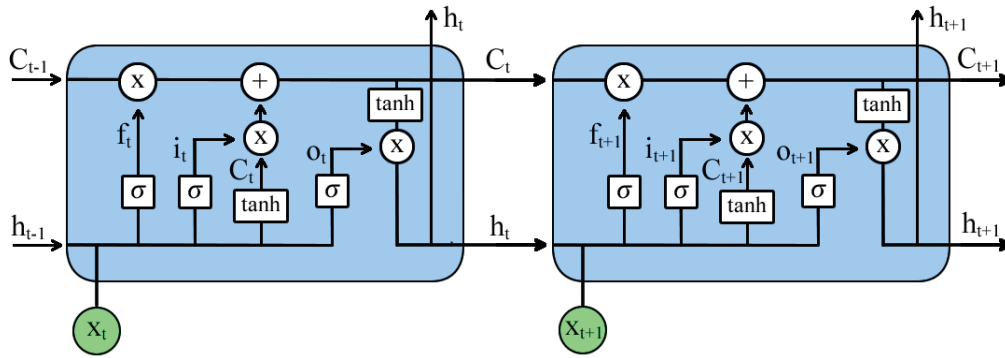
To identify the optimal input data lag and optimal forecasting strategy, an ANN with one hidden layer composed of 40 neurons was used. The number of neurons in the output layer was equal to the number of output variables. Hyperbolic tangent activation function was used in the hidden layer, and linear activation function was used in the output layer. The Adam training algorithm (Kingma and Ba, 2014) was used. Learning rate was set to 0.001, batch size was set to 256 and the number of training epochs was defined using early stopping with maximum training epochs equal to 300 and patience equal to 10 epochs. During hyperparameter tuning phase, other architectures were assessed: one and two hidden layers with 20, 30, 40 and 50 neurons. Batch size was also optimized, testing the following values: 128, 256 and 512. The other hyperparameters were the same ones described above.

### **2.5.2 Deep learning models**

#### **2.5.2.1 Long short-term memory (LSTM)**

LSTM (Hochreiter and Schmidhuber, 1997) is a special type of recurrent neural network (RNN), with advantages over traditional RNN, used to handle sequential data. In this algorithm, when processing input data from time  $t$ , information from previous time steps ( $t-1$ ,  $t-2$ , ...) is considered. An LSTM network is composed of LSTM blocks (Fig. 4). The LSTM block contains forget gate, input gate, output gate, hidden state and cell state. The forget gate decides what information should be removed from the cell state, generating a  $f_t$  value. The input gate decides what information from cell state should be updated, generating an  $i_t$  value. The output gate is responsible to produce  $o_t$ , which is used to compute the hidden state  $h_t$  based on a filtered version of the cell state. More information on LSTM can be found in Lee et al. (2020).





**Fig. 4.** LSTM blocks.  $C_t$  - cell state at time  $t$ ,  $h_t$  - hidden state at time  $t$ ,  $x_t$  - input data at time  $t$ ,  $\sigma$  - sigmoid activation function,  $\tanh$  - hyperbolic tangent activation function,  $f_t$  - forget gate output,  $i_t$  - input gate output,  $o_t$  - output gate output.

To identify the optimal input data lag and optimal forecasting strategy, an LSTM network with two stacked LSTM layers, each one with 30 LSTM units, and two fully connected layers was used. The number of neurons in the first fully connected layers was set to 10. The number of neurons in the second fully connected layer (output layer) was equal to the number of output variables. The hyperparameters of the LSTM layers were kept at their default values. In the first fully connected layer, relu activation function was used, and in the last fully connected layer (output layer), linear activation function was used. The Adam training algorithm (Kingma and Ba, 2014) was used. Learning rate was set to 0.001, batch size was set to 256, and the number of training epochs was defined using early stopping with maximum training epochs equal to 200 and patience equal to 10 epochs.

During hyperparameter tuning phase, different architectures were assessed: one and two LSTM layers with 20, 30 or 40 LSTM units in each one, and none and one fully connected layer with 10 neurons before the output layer. To avoid overfitting, a dropout layer was added after each LSTM layer. The following dropout rates were assessed: 0, 0.2 and 0.4. Batch size was also optimized, testing the following values: 128, 256 and 512. The other hyperparameters were the same ones described above.

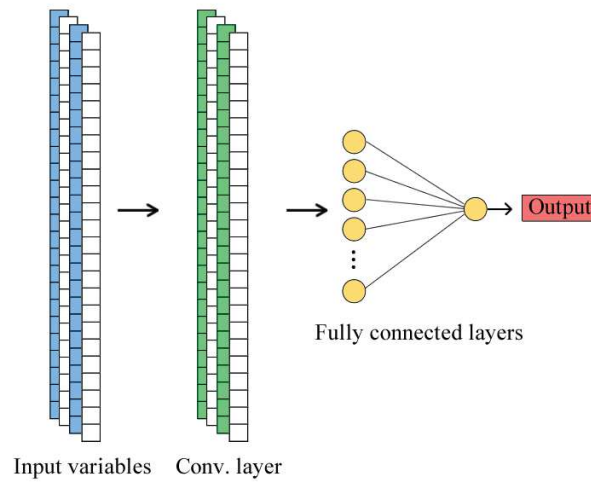
#### 2.5.2.2 Convolutional neural network (CNN)

The main difference between CNN and conventional ANN (i.e., MLP) is the presence of convolution layers. These layers are able to perform automatic feature extraction, capturing features of the input data that are really important to map the relationship between input and output variables. Therefore, CNN can handle raw data without the need for preprocessing or



manual feature extraction. More information on CNN can be found in LeCun et al. (2015) and Goodfellow et al. (2016).

CNN is commonly used for image recognition. In this case, as images have two dimensions, convolutional filters with two dimensions are used. However, when working with sequential data or time series, as in this study, CNN with one-dimensional (1D) convolutional filters (1D CNN) are used (Ferreira and Cunha, 2020). A CNN model with one 1D convolutional layer and two fully connected layers is presented in Fig. 5.



**Fig. 5.** 1D CNN model with one convolutional layer and two fully connected layers.

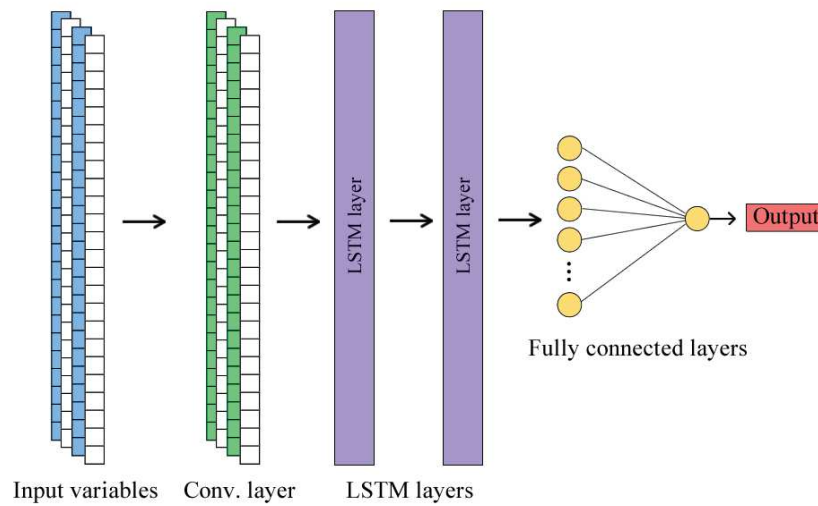
To identify the optimal input data lag and optimal forecasting strategy, a 1D CNN with one convolutional layer and two fully connected layers was used (Fig. 5). In convolutional layer, the following settings were used: 32 filters, kernel size equal to 3, stride equal to 1, padding was set to “same” and relu was used as activation function. The number of neurons in the first fully connected layer was set to 10, and the number of neurons in the second fully connected layer (output layer) was equal to the number of output variables. In the first fully connected layer, relu activation function was used; and in the last fully connected layer, linear activation function was used. Training algorithm, learning rate, batch size and number of training epochs followed the same configurations described for LSTM.

During hyperparameter tuning phase, different architectures were assessed: one and two convolutional layers, each one with 24, 32 or 48 filters and kernel size equal to 2, 3 or 5; none and one pooling layer after convolutional layer(s); and none and one fully connected layer with 10 neurons before the output layer. To avoid overfitting, a dropout layer was added after each convolutional layer. The following dropout rates were assessed: 0, 0.2 and 0.4. Batch size was

also optimized, testing the following values: 128, 256 and 512. The other hyperparameters were the same ones described above.

### 2.5.2.3 CNN-LSTM

A hybrid deep learning model (CNN-LSTM) was also assessed, combining 1D CNN and LSTM. In a CNN-LSTM model, first, one or more convolutional layers are used to extract useful features from the input data and, after, the output is used as input to one or more LSTM layers. Finally, the output from LSTM layers is passed to fully connected layers, in which the results are obtained in the last fully connected layer (output layer). A CNN-LSTM model with one convolutional layer, two LSTM layers and two fully connected layers is presented in Fig. 6.



**Fig. 6.** CNN-LSTM model with one convolutional layer, two LSTM layers and two fully connected layers.

To identify the optimal input data lag and optimal forecasting strategy, a CNN-LSTM with one convolutional layer, two LSTM layers and two fully connected layers was used (Fig. 6). The hyperparameters of convolutional and LSTM layers were the same ones described above for CNN and LSTM models, respectively. The fully connected layers also had the same configurations presented in the description of CNN and LSTM models. Training algorithm, learning rate, batch size and number of training epochs followed the same configurations described for LSTM and CNN.

During hyperparameter tuning phase, different architectures were assessed: one convolutional layer with 24, 32 or 48 filters and kernel size equal to 2, 3 or 5; one and two LSTM layers with 20, 30 or 40 LSTM units in each one; and none and one fully connected layer with 10 neurons before the output layer. To avoid overfitting, a dropout layer was added after

convolutional layer and each LSTM layer. The following dropout rates were assessed: 0, 0.2 and 0.4. Batch size was also optimized, testing the following values: 128, 256 and 512. The other hyperparameters were the same ones described above.

## 2.6 Data normalization

To avoid convergence problems and to remove the effect of different scales, input and output variables were standardized according to the following equation. The mean and standard deviations were obtained using only data from the training subset, not including data from the validation subset and test set. To evaluate the models, ETo was reconverted to its original scale.

$$x_{ni} = \frac{x_i - \mu}{\sigma} \quad (2)$$

where  $x_{ni}$  is the standardized value,  $x_i$  is the observed value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

## 2.7. Performance comparison criteria

The models were evaluated in each test weather station using root mean square error (RMSE), mean absolute error (MAE), coefficient of determination ( $R^2$ ) and Nash-Sutcliffe efficiency coefficient (NSE), according to the following equations. To better evaluate the models, all the metrics were computed considering predicted and observed ETo values individually obtained on each day of the prediction horizon (ind) and considering ETo values accumulated up to each day of the prediction horizon (acc). For RMSE (acc) and MAE (acc), the results were divided by the number of accumulated days in order to keep the unit in mm d<sup>-1</sup>.

$$RMSE = \sqrt{\frac{1}{n} \sum (P_i - O_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum |P_i - O_i| \quad (4)$$

$$R^2 = \left[ \frac{\sum (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{(\sum (P_i - \bar{P})^2)(\sum (O_i - \bar{O})^2)}} \right]^2 \quad (5)$$

$$NSE = 1 - \frac{\sum (P_i - O_i)^2}{\sum (\bar{O} - O_i)^2} \quad (6)$$

where  $P_i$  is the predicted value,  $O_i$  is the observed value,  $\bar{P}$  is the mean of the predicted values,  $\bar{O}$  is the mean of the observed values, and  $n$  is the number of data pairs.

### 3 Results and discussion

#### 3.1 Baselines

Generally, B2 (i.e., baseline based on long-term mean monthly ETo) exhibited better performance than B1 (i.e., baseline based on mean ETo from the last seven days) (Table 3). Comparing mean RMSE values over the prediction horizon and weather stations, B2 showed RMSE (ind) (i.e., RMSE computed considering individual ETo values observed on each day of the prediction horizon,  $\text{mm d}^{-1}$ ) equal to 0.93, while B1 had RMSE (ind) equal to 0.98. Similarly, for RMSE (acc) (i.e., RMSE computed considering ETo values accumulated up to each day of the prediction horizon and also expressed in  $\text{mm d}^{-1}$ ), B2 had RMSE (acc) equal to 0.72, while B1 showed RMSE (acc) equal to 0.75. However, for the first forecasting day, B1 performed better. For the second forecasting day, B1 was similar to B2. From the third day onwards, B2 performed the best. It is important to highlight that B2 had the same RMSE (ind) values for all the days of the prediction horizon since it is based only on long-term mean monthly ETo.

**Table 3.** RMSE values ( $\text{mm d}^{-1}$ ) for ETo forecasted for the prediction horizon in the test weather stations using the proposed baselines.

		Prediction horizon (days)							
Station	Model	1	2	3	4	5	6	7	Mean (SD)
RMSE (ind)									
1	B1	0.93	0.98	1.01	1.02	1.03	1.05	1.06	1.01 (±0.04)
	B2				0.96				0.96 (±0.00)
2	B1	0.81	0.86	0.90	0.93	0.95	0.97	1.00	0.92 (±0.06)
	B2				0.91				0.91 (±0.00)
3	B1	0.92	0.97	0.99	1.00	1.00	1.01	1.00	0.98 (±0.03)
	B2				0.90				0.90 (±0.00)
4	B1	0.90	0.96	0.99	1.01	1.03	1.04	1.05	1.00 (±0.05)
	B2				0.95				0.95 (±0.00)
Mean	B1	0.89	0.94	0.97	0.99	1.00	1.02	1.03	0.98 (±0.05)
	B2				0.93				0.93 (±0.00)
RMSE (acc)									
1	B1	0.93	0.82	0.77	0.73	0.71	0.7	0.69	0.77 (±0.08)
	B2	0.96	0.83	0.75	0.7	0.66	0.63	0.6	0.73 (±0.12)
2	B1	0.81	0.73	0.7	0.68	0.67	0.67	0.67	0.71 (±0.05)
	B2	0.91	0.81	0.76	0.72	0.69	0.67	0.64	0.74 (±0.09)
3	B1	0.92	0.82	0.77	0.74	0.71	0.68	0.67	0.76 (±0.08)
	B2	0.89	0.77	0.69	0.63	0.59	0.55	0.52	0.66 (±0.12)
4	B1	0.9	0.82	0.78	0.75	0.73	0.72	0.71	0.77 (±0.06)
	B2	0.95	0.85	0.78	0.73	0.69	0.65	0.63	0.75 (±0.11)
Mean	B1	0.89	0.80	0.76	0.73	0.71	0.69	0.69	0.75 (±0.07)
	B2	0.93	0.82	0.75	0.70	0.66	0.63	0.60	0.72 (±0.11)

RMSE (ind) - computed considering individual ETo values observed on each day of the prediction horizon,  $\text{mm d}^{-1}$ ; RMSE (acc) - computed considering ETo values accumulated up to each day of the prediction horizon,  $\text{mm d}^{-1}$ .

### 3.2 Traditional machine learning models

The results obtained using the different forecasting strategies (i.e., iterated, direct and MIMO) and the different input data combinations (i.e., (1) lagged ETo, (2) lagged ETo + day

of the year, and (3) lagged ETo + day of the year + additional meteorological data) are presented in Table 4. Regarding the forecasting strategies, direct and MIMO provided the best results, with lower RMSE values in relation to iterated strategy, mainly for RF. Direct and MIMO strategies performed similarly. However, MIMO has a much lower computational cost than direct strategy since it uses only a single model to forecast all the values of the forecasting horizon.

**Table 4.** Mean and standard deviation of RMSE (mm d<sup>-1</sup>) for ETo forecasted for the prediction horizon in the test weather stations using machine learning with iterated, direct and MIMO forecasting strategies in local and regional scenarios, as well as the mean optimal input data lags.

Local				Regional		
Model	Lag	RMSE	RMSE (acc)	Lag	RMSE	RMSE (acc)
Iterated						
RF1	28	0.96 (±0.07)	0.70 (±0.06)	30	0.92 (±0.07)	0.67 (±0.06)
RF2	12	0.95 (±0.06)	0.70 (±0.06)	30	0.94 (±0.06)	0.69 (±0.06)
RF3	9	0.94 (±0.06)	0.68 (±0.06)	5	0.95 (±0.07)	0.69 (±0.05)
ANN1	21	0.93 (±0.06)	0.67 (±0.06)	30	0.91 (±0.06)	0.66 (±0.06)
ANN2	10	0.89 (±0.05)	0.64 (±0.08)	25	0.91 (±0.06)	0.66 (±0.06)
ANN3	3	0.89 (±0.05)	0.64 (±0.07)	4	0.90 (±0.06)	0.65 (±0.07)
Mean	-	0.93 (±0.06)	0.67 (±0.07)	-	0.92 (±0.06)	0.67 (±0.06)
Direct						
RF1	25	0.91 (±0.05)	0.66 (±0.07)	30	0.90 (±0.05)	0.65 (±0.07)
RF2	17	0.91 (±0.05)	0.65 (±0.08)	29	0.91 (±0.06)	0.66 (±0.07)
RF3	12	0.90 (±0.05)	0.65 (±0.07)	22	0.90 (±0.05)	0.65 (±0.07)
ANN1	20	0.92 (±0.06)	0.66 (±0.07)	14	0.91 (±0.06)	0.66 (±0.06)
ANN2	11	0.89 (±0.05)	0.64 (±0.08)	13	0.91 (±0.05)	0.66 (±0.07)
ANN3	4	0.89 (±0.06)	0.63 (±0.07)	7	0.91 (±0.05)	0.65 (±0.07)
Mean	-	0.90 (±0.05)	0.65 (±0.07)	-	0.91 (±0.05)	0.66 (±0.07)
MIMO						
RF1	26	0.90 (±0.05)	0.66 (±0.07)	30	0.90 (±0.06)	0.65 (±0.06)
RF2	20	0.90 (±0.05)	0.66 (±0.07)	30	0.91 (±0.05)	0.67 (±0.06)
RF3	12	0.90 (±0.05)	0.65 (±0.07)	30	0.89 (±0.05)	0.65 (±0.07)
ANN1	19	0.91 (±0.05)	0.66 (±0.07)	15	0.91 (±0.06)	0.66 (±0.06)
ANN2	6	0.88 (±0.04)	0.64 (±0.08)	15	0.90 (±0.05)	0.66 (±0.07)
ANN3	1	0.88 (±0.05)	0.63 (±0.07)	10	0.90 (±0.05)	0.65 (±0.07)
Mean	-	0.90 (±0.05)	0.65 (±0.07)	-	0.90 (±0.05)	0.66 (±0.07)

RMSE (ind) - computed considering individual ETo values observed on each day of the prediction horizon, mm d<sup>-1</sup>; RMSE (acc) - computed considering ETo values accumulated up to each day of the prediction horizon, mm d<sup>-1</sup>.

To better explore the potential of the machine learning models, the models developed with MIMO strategy and optimal input data lag were submitted to hyperparameter tuning (Table 5). In general, after hyperparameter tuning, there were no large performance changes. There were performance improvements only for RF, with slight RMSE reductions. In local scenario, after hyperparameter tuning, the lowest mean RMSE (ind) and mean RMSE (acc) were the same ones observed before hyperparameter tuning, 0.88 and 0.63, respectively. In regional scenario, the lowest mean RMSE (ind) was the same one (0.89), and the lowest mean RMSE (acc) reduced from 0.65 to 0.64.

**Table 5.** RMSE mean values ( $\text{mm d}^{-1}$ ) for ETo forecasted for the prediction horizon in the test weather stations using machine learning models with MIMO forecasting strategy, optimal input data lag and hyperparameter tuning in local and regional scenarios.

RMSE (Local)										
	ind	acc	ind	acc	ind	acc	ind	acc	ind	acc
Model	Station 1		Station 2		Station 3		Station 4		Mean	
RF1	0.93	0.67	0.85	0.63	0.88	0.63	0.92	0.68	0.90	0.65
RF2	0.92	0.66	0.84	0.63	0.88	0.63	0.90	0.65	0.88	0.64
RF3	0.92	0.66	0.83	0.62	0.88	0.62	0.90	0.66	0.88	0.64
ANN1	0.96	0.70	0.85	0.62	0.91	0.65	0.94	0.69	0.91	0.67
ANN2	0.91	0.65	0.82	0.61	0.86	0.60	0.93	0.68	0.88	0.64
ANN3	0.92	0.64	0.82	0.61	0.86	0.60	0.91	0.67	0.88	0.63
RMSE (Regional)										
	ind	acc	ind	acc	ind	acc	ind	acc	ind	acc
Model	Station 1		Station 2		Station 3		Station 4		Mean	
RF1	0.94	0.67	0.83	0.61	0.91	0.66	0.92	0.67	0.90	0.65
RF2	0.93	0.67	0.86	0.64	0.92	0.67	0.92	0.68	0.91	0.66
RF3	0.92	0.65	0.83	0.61	0.88	0.63	0.92	0.68	0.89	0.64
ANN1	0.94	0.68	0.84	0.62	0.92	0.67	0.93	0.69	0.91	0.67
ANN2	0.92	0.66	0.84	0.62	0.94	0.69	0.93	0.68	0.91	0.66
ANN3	0.94	0.68	0.84	0.61	0.90	0.64	0.92	0.67	0.90	0.65



RMSE (ind) - computed considering individual ETo values observed on each day of the prediction horizon, mm d<sup>-1</sup>; RMSE (acc) - computed considering ETo values accumulated up to each day of the prediction horizon, mm d<sup>-1</sup>.

In local scenario, the models developed with lagged ETo + day of the year (RF2 and ANN2) and the models developed with lagged ETo + day of the year + additional meteorological data (RF3 and ANN3) performed slightly better than their versions developed using only lagged ETo (RF1 and ANN1). In regional scenario, for RF models, only RF3 performed better than RF1, and, for ANN models, ANN2 and ANN3 performed better than ANN1. However, it should be highlighted that, in both local and regional scenarios, there were only small performance improvements when additional input data were considered.

Although the best performing model in both local and regional scenarios varied according to the station evaluated, overall, in local scenario, ANN3 performed the best, with mean RMSE (ind) and mean RMSE (acc) equal to 0.88 and 0.63, respectively. In regional scenario, RF3 performed the best, with mean RMSE (ind) and mean RMSE (acc) equal to 0.89 and 0.64, respectively. Although ANN and RF performed better in local and regional scenarios, respectively, the performance differences between the models were small.

When comparing the machine learning models with the best proposed baseline (B2), mean RMSE (ind) reduced from 0.93 to 0.88 (5.4%) and to 0.89 (4.3%), in local and regional scenarios, respectively, and mean RMSE (acc) reduced from 0.72 to 0.63 (12.5%) and to 0.64 (11.1%), in local and regional scenarios, respectively. Although the performance gains found for RMSE (ind) were not large, there were reasonable performance gains for RMSE (acc). It is worth mentioning that, for irrigation scheduling, it can be more important to know ETo accumulated up to a given day of the prediction horizon than individual daily ETo values since daily ETo is accumulated in the soil water balance.

### **3.3 Deep learning models**

For deep learning models, all the forecasting strategies provided similar results in both local and regional scenarios (Table 6). However, in local scenario, direct strategy had a slight advantage over the other forecasting strategies, with lower mean RMSE values.

**Table 6.** Mean and standard deviation of RMSE (mm d<sup>-1</sup>) for ETo forecasted for the prediction horizon in the test weather stations using deep learning with iterated, direct and MIMO forecasting strategies in local and regional scenarios, as well as the mean optimal input data lags.

Local				Regional		
Model	Lag	RMSE	RMSE (acc)	Lag	RMSE	RMSE (acc)
Iterated						
LSTM1	18	0.89 (±0.05)	0.64 (±0.07)	25	0.92 (±0.06)	0.67 (±0.06)
LSTM2	14	0.87 (±0.05)	0.62 (±0.08)	10	0.91 (±0.06)	0.66 (±0.06)
LSTM3	12	0.95 (±0.07)	0.69 (±0.06)	25	0.88 (±0.05)	0.63 (±0.07)
CNN1	22	0.90 (±0.05)	0.65 (±0.07)	20	0.92 (±0.06)	0.67 (±0.06)
CNN2	20	0.89 (±0.05)	0.64 (±0.07)	20	0.90 (±0.05)	0.66 (±0.07)
CNN3	6	0.91 (±0.06)	0.65 (±0.07)	10	0.90 (±0.06)	0.64 (±0.07)
CNN-LSTM1	16	0.89 (±0.05)	0.64 (±0.07)	25	0.90 (±0.06)	0.66 (±0.06)
CNN-LSTM2	16	0.88 (±0.05)	0.63 (±0.07)	30	0.90 (±0.06)	0.65 (±0.07)
CNN-LSTM3	11	0.91 (±0.06)	0.66 (±0.07)	20	0.90 (±0.06)	0.64 (±0.07)
Mean	-	0.90 (±0.05)	0.65 (±0.07)	-	0.90 (±0.06)	0.65 (±0.07)
Direct						
LSTM1	19	0.91 (±0.06)	0.66 (±0.06)	30	0.90 (±0.05)	0.65 (±0.07)
LSTM2	12	0.87 (±0.04)	0.63 (±0.08)	26	0.90 (±0.06)	0.65 (±0.06)
LSTM3	17	0.89 (±0.05)	0.64 (±0.08)	18	0.89 (±0.05)	0.63 (±0.07)
CNN1	19	0.91 (±0.06)	0.65 (±0.07)	27	0.91 (±0.05)	0.66 (±0.07)
CNN2	15	0.89 (±0.05)	0.64 (±0.07)	16	0.91 (±0.05)	0.67 (±0.07)
CNN3	16	0.91 (±0.05)	0.64 (±0.08)	11	0.89 (±0.05)	0.64 (±0.07)
CNN-LSTM1	20	0.89 (±0.06)	0.64 (±0.07)	27	0.90 (±0.06)	0.65 (±0.07)
CNN-LSTM2	15	0.88 (±0.05)	0.63 (±0.08)	25	0.91 (±0.06)	0.66 (±0.07)
CNN-LSTM3	11	0.89 (±0.05)	0.64 (±0.07)	21	0.88 (±0.05)	0.63 (±0.07)
Mean	-	0.89 (±0.05)	0.64 (±0.07)	-	0.90 (±0.05)	0.65 (±0.07)
MIMO						
LSTM1	21	0.89 (±0.05)	0.64 (±0.07)	30	0.90 (±0.05)	0.65 (±0.07)
LSTM2	12	0.88 (±0.04)	0.64 (±0.08)	20	0.90 (±0.05)	0.66 (±0.07)
LSTM3	8	0.92 (±0.05)	0.68 (±0.08)	15	0.89 (±0.05)	0.64 (±0.07)
CNN1	18	0.90 (±0.05)	0.65 (±0.08)	30	0.90 (±0.05)	0.66 (±0.07)

CNN2	12	0.88 ( $\pm 0.05$ )	0.63 ( $\pm 0.08$ )	15	0.90 ( $\pm 0.05$ )	0.66 ( $\pm 0.06$ )
CNN3	8	0.89 ( $\pm 0.04$ )	0.64 ( $\pm 0.08$ )	10	0.88 ( $\pm 0.05$ )	0.64 ( $\pm 0.07$ )
CNN-LSTM1	21	0.89 ( $\pm 0.04$ )	0.64 ( $\pm 0.08$ )	30	0.90 ( $\pm 0.05$ )	0.65 ( $\pm 0.07$ )
CNN-LSTM2	10	0.88 ( $\pm 0.03$ )	0.64 ( $\pm 0.08$ )	20	0.90 ( $\pm 0.05$ )	0.66 ( $\pm 0.07$ )
CNN-LSTM3	15	0.93 ( $\pm 0.03$ )	0.72 ( $\pm 0.09$ )	15	0.89 ( $\pm 0.05$ )	0.64 ( $\pm 0.07$ )
Mean	-	0.90 ( $\pm 0.04$ )	0.65 ( $\pm 0.08$ )	-	0.90 ( $\pm 0.05$ )	0.65 ( $\pm 0.07$ )

RMSE (ind) - computed considering individual ETo values observed on each day of the prediction horizon,  $\text{mm d}^{-1}$ ; RMSE (acc) - computed considering ETo values accumulated up to each day of the prediction horizon,  $\text{mm d}^{-1}$ .

To better explore the potential of the deep learning models, the models developed using MIMO strategy were submitted to hyperparameter tuning with their optimal input data lag (Table 7). MIMO strategy was selected instead of direct and iterated strategies due to the similar results, much lower computational cost in relation to direct strategy, and theoretical advantages over iterated strategy (i.e., it avoids accumulation of prediction errors).

In general, after hyperparameter tuning, there were no large performance changes. In local scenario, the lowest mean RMSE (ind) (considering MIMO strategy) reduced from 0.88 to 0.87 after hyperparameter tuning. The lowest mean RMSE (ind) was the same one (0.63). In regional scenario, after hyperparameter tuning, the lowest mean RMSE (ind) was the same one (0.88), and the lowest mean RMSE (acc) reduced from 0.64 to 0.63.

**Table 7.** RMSE mean values ( $\text{mm d}^{-1}$ ) for ETo forecasted for the prediction horizon in the test weather stations using deep learning models with MIMO forecasting strategy, optimal input data lag and hyperparameter tuning in local and regional scenarios.

Model	RMSE (Local)									
	ind	acc	ind	acc	ind	acc	ind	acc	ind	acc
	Station 1		Station 2		Station 3		Station 4		Mean	
LSTM1	0.94	0.67	0.82	0.61	0.88	0.62	0.92	0.68	0.89	0.64
LSTM2	0.91	0.65	0.82	0.60	0.88	0.62	0.94	0.72	0.89	0.65
LSTM3	0.91	0.64	0.82	0.61	0.89	0.63	0.94	0.71	0.89	0.65
CNN1	0.96	0.69	0.83	0.61	0.91	0.67	0.93	0.69	0.91	0.66
CNN2	0.92	0.65	0.83	0.61	0.87	0.61	0.91	0.67	0.88	0.63
CNN3	0.92	0.64	0.82	0.60	0.88	0.63	0.92	0.70	0.89	0.64

CNN-LSTM1	0.95	0.68	0.83	0.61	0.88	0.62	0.91	0.66	0.89	0.64
CNN-LSTM2	0.90	0.64	0.82	0.61	0.87	0.61	0.89	0.65	0.87	0.63
CNN-LSTM3	0.93	0.66	0.91	0.71	0.89	0.64	0.90	0.66	0.91	0.67
RMSE (Regional)										
	ind	acc	ind	acc	ind	acc	ind	acc	ind	acc
Model	Station 1		Station 2		Station 3		Station 4		Mean	
LSTM1	0.93	0.67	0.83	0.61	0.92	0.67	0.93	0.68	0.90	0.65
LSTM2	0.93	0.67	0.85	0.63	0.92	0.67	0.91	0.67	0.90	0.66
LSTM3	0.93	0.66	0.82	0.60	0.87	0.61	0.91	0.66	0.88	0.63
CNN1	0.93	0.66	0.82	0.60	0.92	0.66	0.93	0.68	0.90	0.65
CNN2	0.92	0.66	0.84	0.62	0.94	0.69	0.93	0.68	0.91	0.66
CNN3	0.94	0.67	0.83	0.61	0.89	0.64	0.91	0.67	0.89	0.65
CNN-LSTM1	0.92	0.66	0.82	0.60	0.92	0.66	0.92	0.68	0.90	0.65
CNN-LSTM2	0.92	0.66	0.84	0.62	0.94	0.69	0.92	0.67	0.90	0.66
CNN-LSTM3	0.92	0.66	0.81	0.59	0.87	0.61	0.90	0.66	0.88	0.63

RMSE (ind) - computed considering individual ETo values observed on each day of the prediction horizon, mm d<sup>-1</sup>; RMSE (acc) - computed considering ETo values accumulated up to each day of the prediction horizon, mm d<sup>-1</sup>.

Overall, as observed for the traditional machine learning models, the different input data combinations did not provide large performance changes. In local scenario, the models developed with lagged ETo + day of the year (input combination 2) generally performed slightly better than the ones developed using only lagged ETo (input combination 1) and lagged ETo + day of the year + additional meteorological data (input combination 3). In regional scenario, the models developed using input combination 3 performed slightly better than their other versions.

In general, in local scenario, the best performance was obtained by the CNN-LSTM2 model, with mean RMSE (ind) and mean RMSE (acc) equal to 0.87 and 0.63, respectively. In regional scenario, LSTM3 and CNN-LSTM3 performed the best, with mean RMSE (ind) and mean RMSE (acc) equal to 0.88 and 0.63, respectively. Although LSTM3 and CNN-LSTM3 obtained the same mean RMSE values over the stations, CNN-LSTM3 had slightly better performance since it showed lower RMSE values in stations 1, 2 and 4.

The combination between CNN and LSTM (i.e., CNN-LSTM models) provided slight performance gains over LSTM and CNN individually used in some cases (Tables 6 and 7). In

addition, as mentioned above, it was the best option to forecast ETo in both local and regional scenarios. However, it also provided worse results than LSTM and CNN individually used in some cases (Tables 6 and 7). Therefore, the results obtained in the present study partially corroborate Barzegar et al. (2020), Huang and Kuo (2018) and Kim and Cho (2019), who reported better performances for CNN-LSTM over CNN and LSTM individually used. It is expected that CNN layers extract features from input data and LSTM layers capture time patterns, improving the model capability and reaching higher performances.

When comparing the best deep learning models with the best traditional machine learning models, the first ones perform only slightly better. In local scenario, mean RMSE (ind) reduced from 0.88 to 0.87 (1.1%), and mean RMSE (acc) was the same one (0.63); in regional scenario, mean RMSE (ind) reduced from 0.89 to 0.88 (1.1%), and mean RMSE (acc) reduced from 0.64 to 0.63 (1.6%).

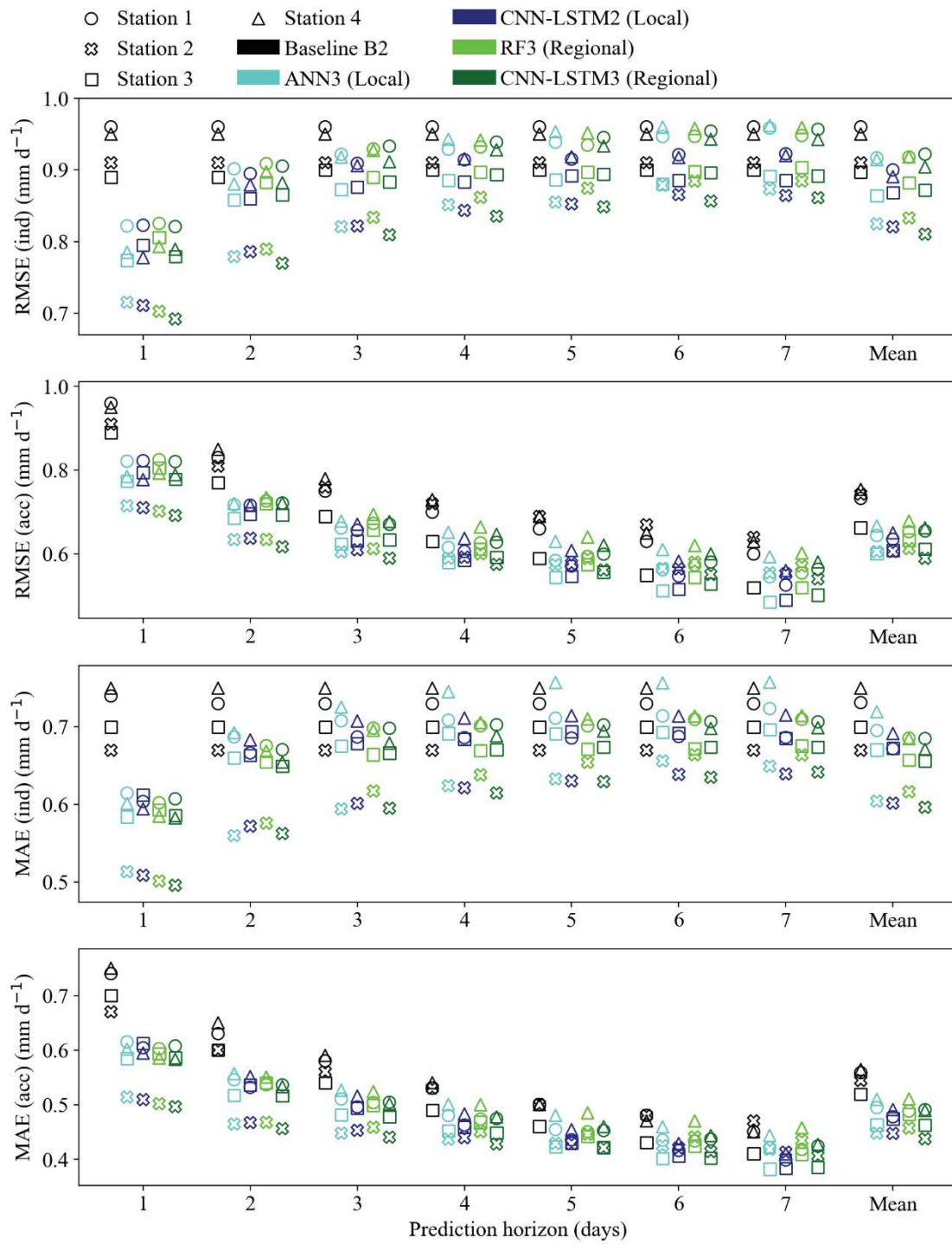
### **3.4 Overall evaluation**

For the traditional machine learning models, direct and MIMO forecasting strategies provided the best results (Table 4). For the deep learning models, all the forecasting strategies performed similarly, with only a slight advantage for direct strategy in local scenario. (Table 6). Although direct strategy has been able to provide good results in all cases, it has the highest computational cost. MIMO provided the best combination between performance and computational cost for the machine learning models and, along with iterated strategy, for the deep learning models. As MIMO provided good results and does not require high computational cost, it can be considered the most promising forecasting strategy assessed. Taieb et al. (2012), comparing five forecasting strategies, recommended MIMO as one of the best options. When forecasting all the prediction horizon at the same time, knowledge learned from one step can benefit others (Ye and Dai, 2019). However, it is important to highlight that, as observed in the present study and typically occurs in machine learning tasks, a given approach does not always perform the best in all cases.

Regarding the different input combinations assessed, they did not provide large performance changes, but when using additional data to lagged ETo (input combinations 2 and 3), there were slight performance gains in some cases (Tables 5 and 7). In addition, the best results for both traditional machine learning and deep learning models were obtained using additional data to lagged ETo. However, in some cases, slightly poorer results were found when additional input data were used. It is also worth mentioning that when using additional input data, the input data lag required generally reduced (Tables 4 and 6).

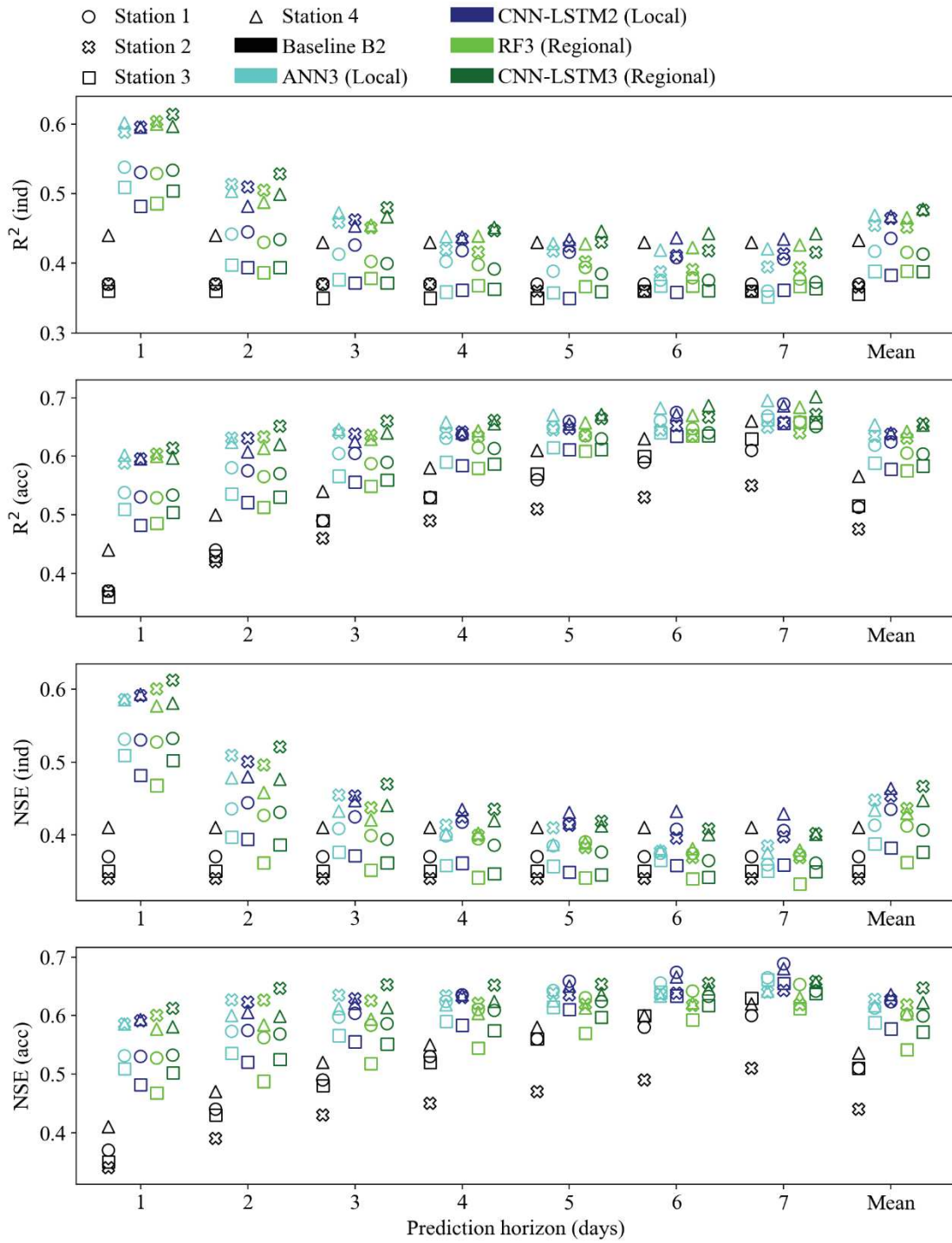
The idea of using multiple time series as input can also be treated as a multivariate time series forecasting task. It is expected that with more input variables, the algorithm will have more information to learn how to estimate the output variable. However, the increase in input data dimensionality can also make the modeling task more complex (Gonzalez-Vidal et al., 2019). Therefore, future studies can explore the potential of feature selection methods to remove irrelevant variables, as done by Feng et al. (2017) and Gonzalez-Vidal et al. (2019).

To better analyze the behavior of the models assessed, the performance of the best local and regional traditional machine learning and deep learning models, as well as the best proposed baseline (B2), throughout the prediction horizon for each weather station is presented in Figs. 7 and 8. To accomplish this, ANN3 and CNN-LSTM2 were considered in local scenario, and RF3 and CNN-LSTM3 were considered in regional scenario.



**Fig. 7.** RMSE and MAE values for each day of the prediction horizon obtained by the best local and regional traditional machine learning and deep learning models, as well as the best proposed baseline for each test weather station. ind - computed considering individual ETo values observed on each day of the prediction horizon; acc - computed considering ETo values accumulated up to each day of the prediction horizon.





**Fig. 8.**  $R^2$  and NSE values for each day of the prediction horizon obtained by the best local and regional traditional machine learning and deep learning models, as well as the best proposed baseline for each test weather station. ind - computed considering individual ETo values observed on each day of the prediction horizon; acc - computed considering ETo values accumulated up to each day of the prediction horizon.



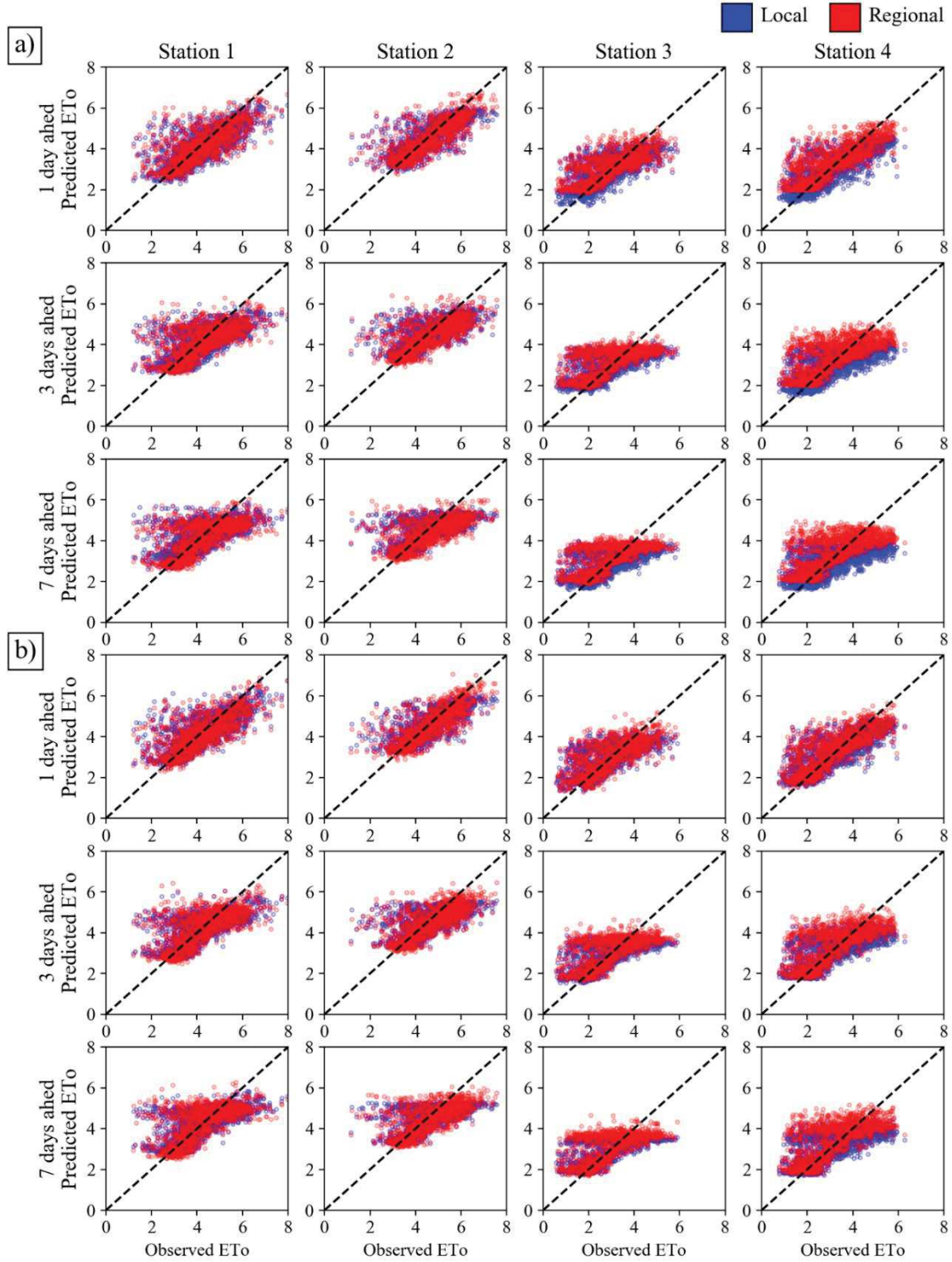
The traditional machine learning and deep learning models exhibited better performances throughout the prediction horizon than the best baseline (B2), generally obtaining lower RMSE and MAE values and higher  $R^2$  and NSE values. Although mean RMSE, MAE,  $R^2$  and NSE (acc) and mean RMSE, MAE,  $R^2$  and NSE (ind) obtained by the machine learning and deep learning models for the whole prediction horizon are not much better than those obtained using B2, there were higher gains on the first and second forecasting days.

The more distant the target day, the higher the forecasting error produced by the models assessed. In general, for the third day onwards, the machine learning and deep learning models provided RMSE, MAE,  $R^2$  and NSE (ind) values similar to those obtained using B2, which is based only on long-term mean monthly ETo. Therefore, it indicates that it is difficult to forecast ETo for large prediction horizons with higher performance than that obtained based only on long-term mean monthly ETo. However, by analyzing RMSE, MAE,  $R^2$  and NSE (acc), it can be noted that they obtained better values with the increase of the number of forecasting days. Although it is more difficult to forecast individual ETo values for more distant days ahead, accumulated ETo values for longer periods are generally more stable, with fewer oscillations, making them easier to forecast. In addition, part of the forecasting errors is probably canceled when daily ETo values are summed up.

The machine learning and deep learning models generally exhibited similar performances throughout the prediction horizon, obtaining similar RMSE, MAE,  $R^2$  and NSE values (Figs. 7 and 8). However, the deep learning models performed slightly better than the traditional machine learning models. With NSE (Fig. 8) as a comparison basis, in local scenario, when using deep learning instead of machine learning, NSE (ind) and NSE (acc) averaged over the stations and prediction horizon increased from 0.42 to 0.43 (2.4%) and from 0.61 to 0.62 (1.6%), respectively. In regional scenario, mean NSE (ind) and mean NSE (acc) increased from 0.41 to 0.42 (2.4%) and from 0.59 to 0.61 (3.4%), respectively. Due to the higher capabilities of deep learning models to handle time series, they are expected to outperform traditional machine learning models. Barzegar et al. (2020), Ferreira and Cunha (2020) and Kim and Cho (2019) also reported better performance of deep learning over traditional machine learning. However, in the present study, the deep learning models provided only slight performance gains. Furthermore, deep learning models generally have more hyperparameters to be adjusted, requiring more time to train them.

The local and regional models exhibited similar performances throughout the prediction horizon (Figs. 7 and 8). It can also be noted in scatter plots presented in Fig. 9. It is important to mention that regional models can be a more flexible approach than local models since they

can be applied in any place of a considered region. In addition, in contrast to B2, all the models developed in this study do not require long-term ETo data. Thus, they can be applied in areas where long-term meteorological data are not available.



**Fig. 9.** Observed vs predicted ETo (mm d<sup>-1</sup>) (1, 3 and 7 days ahead) for the test weather stations using (a) the best local and regional traditional machine learning (ANN3 and RF3) and (b) deep learning (CNN-LSTM2 and CNN-LSTM3) models.

Although the models assessed have high modeling capabilities, especially the deep learning models, and have exhibited better performance than the simple use of long-term mean monthly ETo (i.e., B2), high accuracy in ETo forecasting was not achieved (Figs. 7, 8 and 9). Even when forecasting ETo for only one day ahead, NSE (ind) reached at most 0.61 (station 2 with regional CNN-LSTM3) and NSE (ind) averaged over the stations reached at most 0.56 (using regional CNN-LSTM3). The same behavior was observed for RMSE, MAE and  $R^2$  metrics. The best results were found when estimating ETo accumulated over all the prediction horizon (seven days), in which NSE (acc) reached 0.69 (station 1 using local CNN-LSTM2) and NSE (acc) averaged over the stations reached 0.67 (using local CNN-LSTM2). These results indicate the high complexity of ETo forecasting. Torres et al. (2011), using ANN to forecast ETo up to seven days ahead based on previous ETo (computed using the Hargreaves-Samani equation with only air temperature and  $R_a$ ), found NSE (ind) equal to 0.88 on the first day ahead and 0.77 on the seventh day ahead. Although NSE values reported by Torres et al. (2011) are higher than those found in the present study, in the cited study, ETo was computed using only air temperature and  $R_a$ , which probably facilitates the forecasting task. Landaras et al. (2009), forecasting weekly ETo with ARIMA and ANN, reduced RMSE with respect to weekly historical means only by 6–8%. ETo forecasting is a complex task since ETo is affected by several meteorological variables, which can vary widely from one day to the next.

Another alternative to forecast ETo is the use of weather forecasts. This approach can achieve good accuracy, but it depends on the forecasting performance of the meteorological data used (Cai et al., 2007; Perera et al., 2014; Traore et al., 2017; Yang et al., 2019). A drawback is that although there are public weather forecasts, which may be simpler than non-public ones, some weather forecasts may not be easily and freely available to farmers.

In the present study, although the models developed have not exhibited high accuracies, it is important to highlight that they can be useful tools in places where simpler approaches such as long-term mean monthly ETo are used to forecast ETo. In these situations, they can provide valuable information to improve tasks such as irrigation planning. In addition, the models developed in this study do not depend on external data, requiring only data measured on a local weather station, which facilitates their usage and allows their application in farms without internet connection.

#### 4 Conclusions

This study assesses the potential of deep learning models (LSTM, 1D CNN and CNN-LSTM) and traditional machine learning models (ANN and RF) to forecast multi-step ahead

daily ETo (seven days) using different forecasting strategies. Three input data combinations were assessed. For comparison purposes, two baselines were also used.

Although there were no large performance differences between the forecasting strategies, for the traditional machine learning models, direct and MIMO performed slightly better. For the deep learning models, there was a slight advantage for direct strategy in local scenario. Considering computational cost and performance, MIMO was the best option for the machine learning models, and iterated and MIMO were the best options for the deep learning models.

The deep learning models performed slightly better than the traditional machine learning models, and both were better than the best baseline (B2), mainly on the first and second forecasting days. For the machine learning models, ANN3 (i.e., ANN with input combination 3) and RF3 performed the best in local and regional scenarios, respectively. For the deep learning models, CNN-LSTM2 and CNN-LSTM3 performed the best in local and regional scenarios, respectively. As the performances found in local and regional scenarios were similar, the regional models are recommended due to their higher generalization capacity.

The use of additional input data to lagged ETo (input combinations 2 and 3) provided slight performance improvements in some cases. In addition, the best performing models used additional input data. However, there were only small performance gains, and it did not occur in all the cases assessed. The models developed using additional input data also generally required lower input data lag.

The more distant the target day, the higher the error produced by the models. However, when analyzing accumulated ETo values up to each day of the forecasting horizon, the models obtained better performances with the increase of the forecasting horizon. When forecasting ETo for one day ahead, NSE averaged over the stations reached at most 0.56, and when forecasting ETo accumulated over all the prediction horizon (seven days), it reached 0.67.

Finally, although the models developed have not exhibited high accuracies, they can be useful tools in places where simple approaches such as long-term mean monthly ETo are used to forecast ETo. The models developed can improve tasks such as irrigation planning and do not depend on external data, requiring only data measured on a local weather station. The models are available upon request.

## Acknowledgments

The present study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and CNPq, National Council for Scientific

and Technological Development - Brazil. The authors wish to thank the Brazilian National Institute of Meteorology (INMET) for the meteorological data used.

## References

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and Drainage, Paper No. 56, Food and Agriculture Organization of the United Nations, Rome.
- Alves, W.B., Rolim, G.D.S., Aparecido, L.E.D.O., 2017. Reference evapotranspiration forecasting by artificial neural networks. *Eng. Agric.* 37, 1116–1125. doi:10.1590/1809-4430-eng.agric.v37n6p1116-1125/2017
- Amarasinghe, K., Marino, D.L., Manic, M., 2017. Deep neural networks for energy load forecasting, in: *IEEE International Symposium on Industrial Electronics*. pp. 1483–1488. doi:10.1109/ISIE.2017.8001465
- Ashrafzadeh, A., Kişi, O., Aghelpour, P., Biazar, S.M., Masouleh, M.A., 2020. Comparative Study of Time Series Models, Support Vector Machines, and GMDH in Forecasting Long-Term Evapotranspiration Rates in Northern Iran. *J. Irrig. Drain. Eng.* 146. doi:10.1061/(ASCE)IR.1943-4774.0001471
- Bachour, R., Maslova, I., Ticlavilca, A.M., Walker, W.R., McKee, M., 2016. Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. *Stoch. Environ. Res. Risk Assess.* 30, 103–117. doi:10.1007/s00477-015-1039-z
- Barzegar, R., Aalami, M.T., Adamowski, J., 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* 34, 415–433. doi:10.1007/s00477-020-01776-2
- Ben Taieb, S., Bontempi, G., Atiya, A.F., Sorjamaa, A., 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Syst. Appl.* 39, 7067–7083. doi:10.1016/j.eswa.2012.01.039
- Ben Taieb, S., Sorjamaa, A., Bontempi, G., 2010. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing* 73, 1950–1957. doi:10.1016/j.neucom.2009.11.030
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324
- Cai, J., Liu, Y., Lei, T., Pereira, L.S., 2007. Estimating reference evapotranspiration with the FAO Penman-Monteith equation using daily weather forecast messages. *Agric. For. Meteorol.* 145, 22–35. doi:10.1016/j.agrformet.2007.04.012



- Feng, C., Cui, M., Hodge, B.M., Zhang, J., 2017. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl. Energy* 190, 1245–1257. doi:10.1016/j.apenergy.2017.01.043
- 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. doi:10.1016/j.agwat.2020.106113
- Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. *J. Hydrol.* 572, 556–570. doi:10.1016/j.jhydrol.2019.03.028
- Gao, S., Wang, X., Miao, X., Su, C., Li, Y., 2019. ASM1D-GAN: An Intelligent Fault Diagnosis Method Based on Assembled 1D Convolutional Neural Network and Generative Adversarial Networks. *J. Signal Process. Syst.* 91, 1237–1247. doi:10.1007/s11265-019-01463-8
- González-Vidal, A., Jiménez, F., Gómez-Skarmeta, A.F., 2019. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy Build.* 196, 71–82. doi:10.1016/j.enbuild.2019.05.021
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. MIT press.
- Haidar, A., Verma, B., 2018. Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network. *IEEE Access* 6, 69053–69063. doi:10.1109/ACCESS.2018.2880044
- Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Comput.* 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Huang, C.J., Kuo, P.H., 2018. A deep cnn-lstm model for particulate matter (Pm2.5) forecasting in smart cities. *Sensors (Switzerland)* 18. doi:10.3390/s18072220
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* 574, 1029–1041. doi:10.1016/j.jhydrol.2019.04.085
- Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* doi:10.1016/j.compag.2018.02.016
- Karbasi, M., 2018. Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet- Gaussian Process Regression Model. *Water Resour. Manag.* 32, 1035–1052. doi:10.1007/s11269-017-1853-9
- Kim, T.Y., Cho, S.B., 2019. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* 182, 72–81. doi:10.1016/j.energy.2019.05.230

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

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

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature*. doi:10.1038/nature14539

Lee, T., Shin, J.Y., Kim, J.S., Singh, V.P., 2020. Stochastic simulation on reproducing long-term memory of hydroclimatological variables using deep learning model. *J. Hydrol.* 582. doi:10.1016/j.jhydrol.2019.124540

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

Pereira, L.S., Allen, R.G., Smith, M., Raes, D., 2015. Crop evapotranspiration estimation with FAO56: Past and future. *Agric. Water Manag.* doi:10.1016/j.agwat.2014.07.031

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

Saggi, M.K., Jain, S., 2019. Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. *Comput. Electron. Agric.* 156, 387–398. doi:10.1016/j.compag.2018.11.031

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., Jung, J., 2020. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks* 121, 396–408. doi:10.1016/j.neunet.2019.09.033

Shen, C., 2018. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resour. Res.* doi:10.1029/2018WR022643

Silva Júnior, J.C., Medeiros, V., Garrozi, C., Montenegro, A., Gonçalves, G.E., 2019. Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. *Comput. Electron. Agric.* 166, 105017. doi:10.1016/j.compag.2019.105017

Son, H., Kim, C., 2020. A Deep Learning Approach to Forecasting Monthly Demand for Residential–Sector Electricity. *Sustainability* 12, 3103. doi:10.3390/su12083103

Tian, C., Ma, J., Zhang, C., Zhan, P., 2018. A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies* 11. doi:10.3390/en11123493

- Tian, L., Leason, Z.T., Quiring, S.M., 2020. Developing a hybrid drought index: Precipitation Evapotranspiration Difference Condition Index. *Clim. Risk Manag.* 29, 100238. doi:10.1016/j.crm.2020.100238
- Torres, A.F., Walker, W.R., McKee, M., 2011. Forecasting daily potential evapotranspiration using machine learning and limited climatic data. *Agric. Water Manag.* 98, 553–562. doi:10.1016/j.agwat.2010.10.012
- Trajkovic, S., Todorovic, B., Stankovic, M., 2003. Forecasting of reference evapotranspiration by artificial neural networks. *J. Irrig. Drain. Eng.* 129, 454–457. doi:10.1061/(ASCE)0733-9437(2003)129:6(454)
- Traore, S., Luo, Y., Fipps, G., 2017. Gene-Expression Programming for Short-Term Forecasting of Daily Reference Evapotranspiration Using Public Weather Forecast Information. *Water Resour. Manag.* 31, 4891–4908. doi:10.1007/s11269-017-1784-5
- Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water (Switzerland)*. doi:10.3390/w11050910
- UNEP (United Nations Environmental Programme). 1997. *World Atlas of Desertification*. Editorial commentary by N. Middleton and D.S.G. Thomas, London, Edward Arnold.
- Yang, Y., Luo, Y., Wu, C., Zheng, H., Zhang, L., Cui, Y., Sun, N., Wang, L., 2019. Evaluation of six equations for daily reference evapotranspiration estimating using public weather forecast message for different climate regions across China. *Agric. Water Manag.* 222, 386–399. doi:10.1016/j.agwat.2019.06.014
- Ye, R., Dai, Q., 2019. MultiTL-KELM: A multi-task learning algorithm for multi-step-ahead time series prediction. *Appl. Soft Comput. J.* 79, 227–253. doi:10.1016/j.asoc.2019.03.039
- Zhou, Y., Chang, F.J., Chang, L.C., Kao, I.F., Wang, Y.S., 2019. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J. Clean. Prod.* 209, 134–145. doi:10.1016/j.jclepro.2018.10.243



### **Article 3: Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil**

**Abstract:** Reference evapotranspiration (ET<sub>o</sub>) can be estimated using the FAO56-Penman-Monteith (FAO56-PM) equation but it requires commonly unavailable meteorological data. Therefore, this study assessed different approaches to estimate ET<sub>o</sub> based on temperature and relative humidity, and temperature only across Brazil, as follows: (i) using the FAO56-PM equation with missing data estimated based on FAO56 methodologies; (ii) using the FAO56-PM equation with missing data estimated based on machine learning; and (iii) estimating ET<sub>o</sub> directly using machine learning. The FAO56-PM equation was also calibrated through linear regression and by calibrating the methodologies used to estimate missing data. The potential benefits of using multi-task learning (MTL) and clustering were also investigated. Data from 437 weather stations were used. Artificial neural network (ANN), random forest (RF), extreme gradient boosting (XGBoost) and multivariate adaptive regression splines (MARS) were employed. In both general and clustering scenarios, calibrating the FAO56-PM equation using linear regression provided slightly better results than calibrating the methodologies used to estimate missing data. In contrast to temperature- and relative humidity-based FAO56-PM equation, its temperature-based version performed better before both calibration types assessed. The machine learning models performed the best to estimate ET<sub>o</sub> and missing data. Combining the machine learning models with the FAO56-PM equation to estimate ET<sub>o</sub> performed similarly to using them individually. MTL and single-task learning (STL) provided similar results. In the general scenario, for the temperature-based models, using PM-ANN-STL increased mean NSE from 0.49 to 0.53 in relation to the non-calibrated FAO56-PM equation. For the temperature- and relative humidity-based models, using ANN and RF developed with STL or MTL increased NSE from 0.56 to 0.67 in relation to the FAO56-PM equation calibrated using linear regression. When using the clustering strategy, performance gains were obtained in estimating ET<sub>o</sub> with the temperature-based models, increasing mean NSE up to 0.58.

**Keywords:** generalizability, MARS, neural networks, solar radiation, wind speed

## **1 Introduction**

Knowing evapotranspiration (ET) is essential in hydrological studies, agricultural planning, water resource management and irrigation scheduling. Due to the difficulty in measuring ET directly, it is commonly obtained based on reference evapotranspiration (ET<sub>o</sub>) and crop coefficient (K<sub>c</sub>) (Pereira et al., 2015). ET<sub>o</sub> can be estimated using the FAO-56

Penman-Monteith (FAO56-PM) equation. However, to use this equation, air temperature, relative humidity, solar radiation and wind speed measurements are required, which are not fully available in some regions. Among the aforementioned data, air temperature is the most commonly and easily measured, which justifies the study of ETo models based on this variable (Mattar et al., 2016; Paredes and Pereira, 2019). In addition to temperature, the use of relative humidity promotes a better estimation of ETo at a low additional cost (Ferreira et al., 2019; Exner-Kittridge, 2012; Valiantzas, 2012; Valiantzas, 2018a).

To estimate ETo under limited meteorological data availability, several empirical equations with fewer data requirements have been studied (Ahmadi and Javanbakht, 2020; Awal et al., 2020; Cunha et al., 2017; Djaman et al., 2018; Hadria et al., 2021; Valiantzas, 2013). In addition to alternative equations, the FAO56-PM equation can be used by estimating missing data using methodologies proposed in the FAO56 bulletin (Allen et al., 1998; Paredes and Pereira, 2019). Overall, this approach has shown good performance in Brazil and other parts of the world. Ferreira et al. (2019), in a study carried out in Brazil, reported that the FAO56-PM equation with missing data performed better than most of the empirical equations assessed. Almorox et al. (2018) found better results for the FAO56-PM equation using only measured data on temperature in relation to the well-known Hargreaves-Samani equation. Similar results were also reported by Alencar et al. (2015) and Ren et al. (2016). These results are possibly related to the physical basis of the FAO56-PM equation, which is partially conserved when estimating missing data. However, it is fair to mention that the better performance of this equation may be related to the fact that the FAO56-PM equation with full data set is commonly used as reference for the assessment of ETo models.

In addition to empirical equations, several studies have employed machine learning models to estimate ETo (Ferreira et al., 2019; Huang et al., 2019; Kisi and Alizamir, 2018; Mehdizadeh et al., 2017; Yu et al., 2020). Due to their superior modeling capabilities, machine learning models generally show better performances than traditional empirical equations. Other alternatives to address meteorological data scarcity include the use of remote sensing data (De Bruin et al., 2016) and reanalysis data (Paredes et al., 2018b).

In Brazil, studies using machine learning to estimate ETo are generally restricted to parts of the country (Althoff et al., 2018; Ferreira et al., 2021; Reis et al., 2019; Tangune and Escobedo, 2018). According to our knowledge, only Ferreira et al. (2019) used machine learning models to estimate ETo based on limited meteorological data considering the entirety of Brazil. This study evaluated artificial neural network (ANN) and support vector machine (SVM), reporting better results for the first model. Brazil has a large territorial area, 8.516

million km<sup>2</sup>, with several climate types. Thus, studies that explore all the Brazilian territory are encouraged, since the results tend to have high representativeness. In addition, according to the National Water and Sanitation Agency of Brazil (ANA, 2021), Brazil has an irrigated area of 8.2 million hectares, with a prospect of incorporating 4.2 million hectares by 2040, which reinforces the need to develop reliable ETo models, even under data scarcity conditions. Therefore, the assessment of state-of-the-art machine learning models deserves special attention.

An important issue for ETo models is their spatial generalization, which refers to the ability of a model to perform well outside its development place (Ferreira et al., 2021; Kiafar et al., 2017). Many studies fail to address it, since they assess the models using only data of the same weather station where the models were developed. To improve the generalization of machine learning models, regional models have a key role. In contrast to local models, which are developed using data from a single weather station, regional models are developed using pooled data from several stations. Thus, they can be applied in any place of a particular region, making them more useful. However, most studies have applied only local models. In Brazil, only a limited number of studies have investigated the potential of regional models (Ferreira et al., 2019; Ferreira et al., 2021; Reis et al., 2019; Zanetti et al., 2019). On the other hand, when developing regional models for large areas, such as the whole of Brazil, the high climatic variability can lead to poorer performances, given the presence of different patterns between input variables and ETo within the same region. In this context, the use of clustering algorithms can guide the definition of smaller regions with homogeneous climatic conditions (clusters). Thus, specific models can be developed for each region, which can contribute to achieving better results, as reported by Ferreira et al. (2019).

Previous studies have been limited to using machine learning models to estimate ETo directly from the available data. However, it is also possible to use machine learning models to estimate missing data, which can be used to estimate ETo with the FAO56-PM equation. In this case, the physical basis of the FAO56-PM equation could be used to improve the performance of the estimation. In addition, it can also make the ETo estimation process clearer in relation to the estimation performed using only machine learning models. Another advantage is that the meteorological data estimated can be used for other purposes. Despite these possible benefits, according to our knowledge, no study has investigated this strategy.

Another approach that could improve the performance of machine learning models for the estimation of ETo is multi-task learning (MTL) (Caruana, 1997; Ng et al., 2019; Nunes et al., 2019; Padarian et al., 2019). In MTL, multiple related variables are predicted

simultaneously. Thus, the information shared among tasks can improve the overall performance in relation to individual predictions (single-task learning (STL)). In this sense, to take advantage of the potential of MTL, the ETo estimation task should be reframed, estimating ETo and related meteorological variables, such as solar radiation and relative humidity, together. Therefore, the information learned by a machine learning model to estimate the related variables could improve the estimation of ETo and vice versa. MTL can be used focusing on the estimation of ETo or the estimation of missing data for use in the FAO56-PM equation. According to our knowledge, no study has investigated the use of MTL to estimate ETo.

The present study aims to assess different approaches to estimate ETo based on temperature and relative humidity, and temperature only across Brazil, as follows: (i) using the FAO56-PM equation with missing data estimated based on methodologies proposed in FAO56; (ii) using the FAO56-PM equation with missing data estimated based on machine learning models; and (iii) estimating ETo directly using machine learning models. In objective (i), two additional strategies are evaluated independently: the calibration of the methodologies proposed in FAO56 to estimate missing data, and the calibration of the FAO56-PM equation by means of linear regression. We also investigated the potential of MTL and clustering to improve the estimation of ETo and missing data.

## **2 Materials and methods**

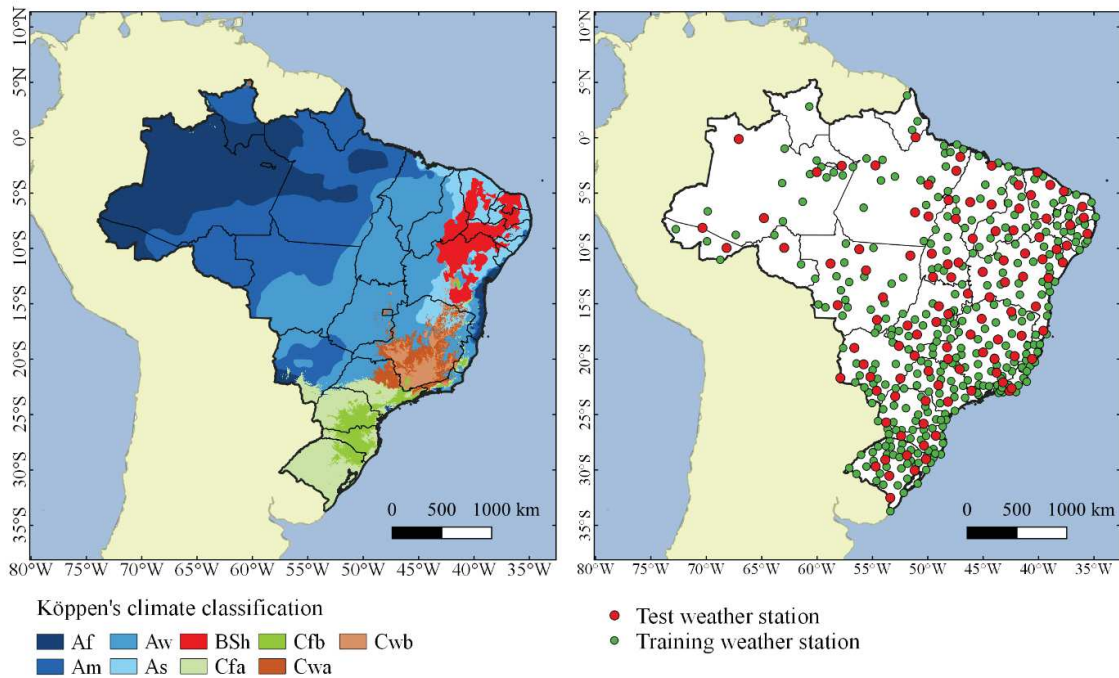
### **2.1 Database and study sites**

Hourly data from 437 weather stations of the Brazilian National Institute of Meteorology (INMET), distributed across Brazil, were used. The period from 2004-2018 was considered. However, the data length varied among stations due to their different epochs of creation, with a mean length of  $10.5 \pm 2.6$  years. Data on maximum and minimum air temperature, maximum and minimum relative humidity, solar radiation and wind speed were obtained. Hourly data were transformed to a daily scale. Wind speed, measured at 10 m height, was converted to 2 m based on recommendations of Allen et al. (1998). Daily records with missing or faulty data were removed. Faulty data were identified when at least one of the following conditions was met: minimum temperature higher than maximum temperature; relative humidity out of the range 0-100%; minimum relative humidity higher than maximum relative humidity; negative wind speed; negative solar radiation; and solar radiation higher than extraterrestrial radiation.

The mean values and standard deviations over the entire dataset were  $29.5 \pm 5.0^\circ\text{C}$  for maximum temperature,  $18.6 \pm 4.5^\circ\text{C}$  for minimum temperature,  $89.9 \pm 9.4\%$  for maximum relative humidity,  $46.6 \pm 16.8\%$  for minimum relative humidity,  $1.92 \pm 0.52$  kPa for actual vapor

pressure (computed using maximum and minimum relative humidity),  $18.2 \pm 6.1 \text{ MJ m}^{-2} \text{ d}^{-1}$  for solar radiation,  $1.53 \pm 0.94 \text{ m s}^{-1}$  for wind speed, and  $4.13 \pm 1.54 \text{ mm d}^{-1}$  for ETo.

To develop and test the proposed models, the dataset was split into training and test datasets. The first one was used to develop the machine learning models and to calibrate the FAO56-PM equation (with missing data) through linear regression and through the calibration of the FAO methodologies used to estimate missing data. The second dataset was used to assess the performance of the models studied. To make the performance assessment more robust, we considered data splitting in both time and space. Thus, the training dataset was composed of 330 weather stations with data up to 2013, and the test dataset was composed of 107 stations with data from 2014-2018. The weather stations splitting and the Köppen's climatic classification (Alvares et al., 2013) for the study area are presented in Fig. 1. Test stations were manually selected in order to provide a good distribution over the Brazilian territory.



**Fig. 1.** Location and training/test splitting of the weather stations and climate classification of the study area. A - Tropical zone; Af - without dry season; Am - monsoon; Aw - with dry winter; As - with dry summer; B - Dry zone; BS - semi-arid; BSh - low latitude and altitude; C - Humid subtropical zone; Cf - oceanic climate, without dry season; Cfa - with hot summer; Cfb - with temperate summer; Cw - with dry winter; Cwa - and hot summer; Cwb - and temperate summer.

To produce models with high generalization ability, the models were developed/calibrated using pooled data from all the stations of the training dataset. In addition,

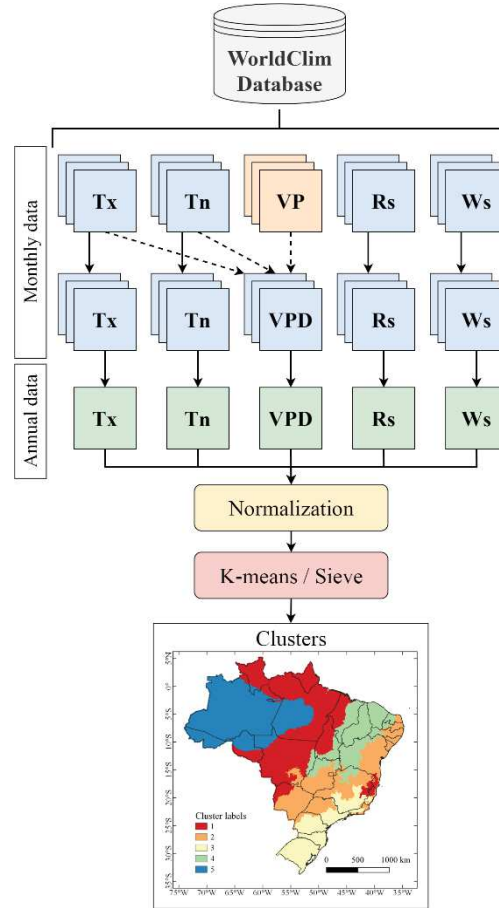
after identifying the most promising models, they were developed for specific regions of Brazil (clusters) defined based on K-means and data from WorldClim (Fick and Hijmans, 2017). Thus, each set of training and test stations contained in each region (cluster) was independently used to develop and test the models. The performance assessment of the models was performed individually for each weather station of the test dataset.

To divide Brazil into climatic homogeneous regions, gridded mean monthly historical data (1970-2000) from WorldClim dataset version 2.1 (Fick and Hijmans, 2017) were used. Data were downloaded with a spatial resolution of 5 minutes. Maximum and minimum air temperature, water vapor pressure, solar radiation and wind speed data were collected. Monthly vapor pressure deficit, which was used to replace water vapor pressure, was computed from maximum and minimum air temperature and water vapor pressure, as suggested by Allen et al. (1998). Mean annual data were computed from monthly data. After that, data were normalized according to Eq. 1. Then, the mentioned data (i.e, maximum and minimum air temperature, vapor pressure deficit, solar radiation and wind speed) were used as input for K-means to define five clusters with similar characteristics (climatic homogeneous regions). The mentioned variables were selected because they are the base variables used in the FAO56-PM equation to estimate ETo. The number of clusters was defined with support of the Elbow Method (Kodinariya and Makwana, 2013). After defining clusters, the sieve tool of the rasterio library (Python) was used to identify small isolated areas (pixels) and replaces their values with the pixel value of the largest neighbor cluster. The procedure followed to define clusters is also presented in Fig. 2.

$$x_{ni} = \frac{x_i - \mu}{\sigma} \quad (1)$$

where  $x_{ni}$  is the standardized value,  $x_i$  is the observed value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.





**Fig. 2.** Flowchart of the procedure adopted to divide Brazil into clusters (climatic homogeneous regions) using K-means and data from WorldClim dataset. Tx - maximum temperature; Tn - minimum temperature; VP - water vapor pressure; VPD - vapor pressure deficit; Rs - solar radiation; Ws - wind speed.

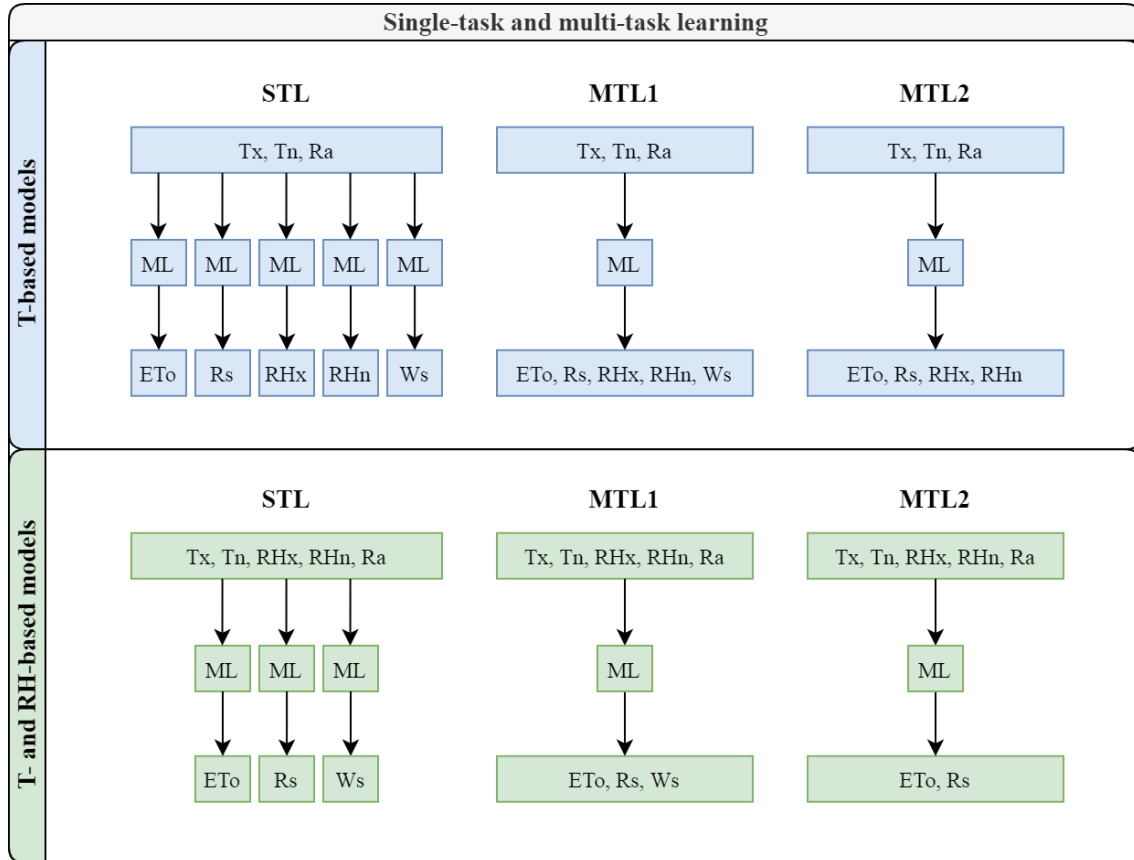
## 2.2 Proposed modeling framework

Different approaches for estimating ETo based on temperature and relative humidity and based on temperature only were assessed. The first approach was to use the FAO56-PM equation with missing data estimated based on traditional methodologies presented in the FAO bulletin (Allen et al., 1998). The equations used to estimate missing data were considered in their original and calibrated forms. In addition, a calibration of the FAO56-PM equation using simple linear regression, as presented in Allen et al. (1998), was performed. In this case, the equations used to estimate missing data were used in their original forms. The second approach was similar to the first one, but using machine learning models to estimate missing data. The third approach was to estimate ETo from limited data directly using machine learning models. All the approaches were developed/calibrated considering Brazil as whole (general scenario) and, for the first approach (i.e., FAO56-PM equation with missing data estimated based on

traditional methodologies) and for the most promising models of the other approaches, also considering individual clusters defined using K-means (clustering scenario).

To explore the potential of multi-task learning (MTL), auxiliary target variables were added when estimating ETo and missing data. Two MTL strategies were considered: MTL1, models used ETo, solar radiation, maximum and minimum relative humidity and wind speed as output variables, in the case of temperature-based models, and these same variables except for maximum and minimum relative humidity, in the case of temperature- and relative humidity-based models; MTL2, models used the same output variables considered in MTL1, except for wind speed. Proposing these two MTL options is because wind speed has a weak relation with the input variables considered in the present study (temperature, relative humidity and extraterrestrial radiation). Thus, trying to add it as an auxiliary target variable may decrease the performance of the models. Machine learning models based on single-task learning (STL) were also assessed by developing individual models to estimate each variable of interest. All the machine learning models developed were based on two types of input data: temperature-based, which considers maximum and minimum temperature and extraterrestrial solar radiation as input, and temperature- and relative humidity-based, which consider also maximum and minimum relative humidity. An overview of the use of STL and MTL to develop the machine learning models is shown in Fig. 3. The following machine learning models were used: artificial neural network (ANN), random forest (RF), extreme gradient boosting (XGBoost) and multivariate adaptive regression splines (MARS). As MTL requires models that support multi-output, only ANN and RF were used with MTL.





**Fig. 3.** An overview of the use of single-task learning (STL) and multi-task learning (MTL) to develop machine learning models to estimate ETo and missing data. Tx - maximum temperature; Tn - minimum temperature; RHx - maximum relative humidity; RHn - minimum relative humidity; Ra - extraterrestrial radiation; ML - machine learning model; ETo - reference evapotranspiration, Rs - solar radiation, Ws - wind speed.

### 2.3 Reference model (FAO56 Penman-Monteith with full data set)

To develop/calibrate and assess the performance of the models, ETo computed using the FAO56-PM equation (Eq. 2) with full data set was used as reference. All calculation procedures were performed following the FAO56 bulletin (Allen et al., 1998).

$$ET_o = \frac{0.408 \Delta (R_n - G) + \gamma \frac{900}{T_{avg} + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 u_2)} \quad (2)$$

where  $ET_o$  is the reference evapotranspiration ( $\text{mm d}^{-1}$ ),  $R_n$  is the net solar radiation ( $\text{MJ m}^{-2} \text{d}^{-1}$ ),  $G$  is the soil heat flux ( $\text{MJ m}^{-2} \text{d}^{-1}$ ) (considered to be null for daily estimates),  $T_{avg}$  is the daily mean air temperature ( $^{\circ}\text{C}$ ),  $u_2$  is the wind speed at a 2 m height ( $\text{m s}^{-1}$ ),  $e_s$  is the saturation vapor pressure (kPa),  $e_a$  is the actual vapor pressure (kPa) (obtained using maximum and minimum relative humidity),  $\Delta$  is the slope of the saturation vapor pressure function ( $\text{kPa } ^{\circ}\text{C}^{-1}$ ), and  $\gamma$  is the psychrometric constant ( $\text{kPa } ^{\circ}\text{C}^{-1}$ ).

### 2.4 FAO56 Penman-Monteith with missing data

To estimate ETo under limited data availability using the FAO56-PM equation, actual vapor pressure was estimated using Eq. 3 (in case of missing data on relative humidity), and solar radiation was estimated using Eq. 4, as suggested in the FAO56 bulletin (Allen et al., 1998). To estimate missing wind speed, in addition to considering the world average wind speed value ( $2.0 \text{ m s}^{-1}$ ) presented in the FAO56 bulletin, average wind speed over the training dataset was considered. Mean wind speed was obtained for Brazil as a whole (general scenario) and for individual clusters (clustering scenario). Maximum and minimum relative humidity (which are used to compute actual vapor pressure), solar radiation, and wind speed were also estimated using machine learning models, as described later.

$$e_a = e^\circ(T_{\text{dew}}) = 0.611 \exp \left[ \frac{17.27 (T_{\text{min}} - a_T)}{(T_{\text{min}} - a_T) + 237.3} \right] \quad (3)$$

where  $e_a$  is the actual vapor pressure (kPa),  $e^\circ$  is the saturation vapor pressure (kPa),  $T_{\text{min}}$  is the minimum air temperature ( $^\circ\text{C}$ ), and  $a_T$  is a correction factor with a common range of  $0\text{-}4^\circ\text{C}$  (Paredes and Pereira, 2019).

$$R_s = k_{R_s} R_a (T_{\text{max}} - T_{\text{min}})^{0.5} \quad (4)$$

where  $R_s$  is the solar radiation ( $\text{MJ m}^{-2} \text{ day}^{-1}$ ),  $k_{R_s}$  is an empirical radiation adjustment coefficient with a common range of  $0.16\text{-}0.19 \text{ } (^\circ\text{C}^{-0.5})$  (Allen et al., 1998);  $R_a$  is the extraterrestrial radiation ( $\text{MJ m}^{-2} \text{ day}^{-1}$ ),  $T_{\text{max}}$  is the maximum air temperature ( $^\circ\text{C}$ ), and  $T_{\text{min}}$  is the minimum air temperature ( $^\circ\text{C}$ ).

When estimating actual vapor pressure from temperature, minimum temperature can be considered as an approximation of dew point temperature. However, this condition is not always met. Thus, to better estimate actual vapor pressure, a correction factor  $a_T$  can be subtracted from minimum temperature (Allen et al., 1998; Paredes and Pereira, 2019). Mean temperature can also be used for humid regions (Paredes and Pereira, 2019). However, according to our previous tests, for the conditions of the present study, using minimum temperature provided better results than using mean temperature (both alternatives were evaluated adopting an optimal correction factor). Therefore, the direct use of minimum temperature ( $a_T=0$ ) and the use of minimum temperature corrected ( $a_T \neq 0$ ) were considered in the present study. To define the optimal value of  $a_T$ , an optimization algorithm was used to find the  $a_T$  value that minimizes the estimation error (mean squared error) of actual vapor pressure. Actual vapor pressure computed using maximum and minimum relative humidity was used as reference. The optimization process was performed using data from the training set. The optimization algorithm used was the function “minimize” from the SciPy library for Python.

The optimal  $a_T$  value was defined for Brazil as a whole (general scenario) and for individual clusters (clustering scenario).

To estimate solar radiation, the default value (0.16) of the empirical radiation adjustment coefficient ( $k_{Rs}$ ) and a calibrated value of  $k_{Rs}$  were considered. To calibrate  $k_{Rs}$ , a procedure similar to the one described to calibrate  $a_T$  in the estimation of actual vapor pressure was followed. Thus,  $k_{Rs}$  value was defined in order to minimize the estimation error of solar radiation, using measured solar radiation as reference. The optimal  $k_{Rs}$  value was defined for Brazil as a whole (general scenario) and for individual clusters (clustering scenario).

The FAO56-PM equation with missing data estimated using the original FAO methodologies (i.e., with default parameters) was also calibrated using simple linear regression. This type of calibration is suggested by Allen et al. (1998) and is very common in studies addressing empirical equations (Ferreira et al., 2019; Feng et al., 2017; Zanetti et al., 2019). To perform this calibration,  $ET_o$  computed using the FAO56-PM equation with full data set was used as target variable ( $y$ ) and  $ET_o$  computed using the FAO56-PM equation with missing data was used as input variable ( $x$ ). Thus, calibrated  $ET_o$  is obtained as shown in Eq. 5. This calibration was performed for Brazil as a whole (general scenario) and for individual clusters (clustering scenario).

$$ET_{o\text{ cal}} = a + b(ET_o) \quad (5)$$

where  $ET_{o\text{ cal}}$  is the calibrated reference evapotranspiration ( $\text{mm d}^{-1}$ ),  $a$  and  $b$  are the calibration parameters, and  $ET_o$  is the reference evapotranspiration estimated using the equation to be calibrated (FAO56-PM with missing data) ( $\text{mm d}^{-1}$ ).

## 2.5 Machine learning models and multi-task learning

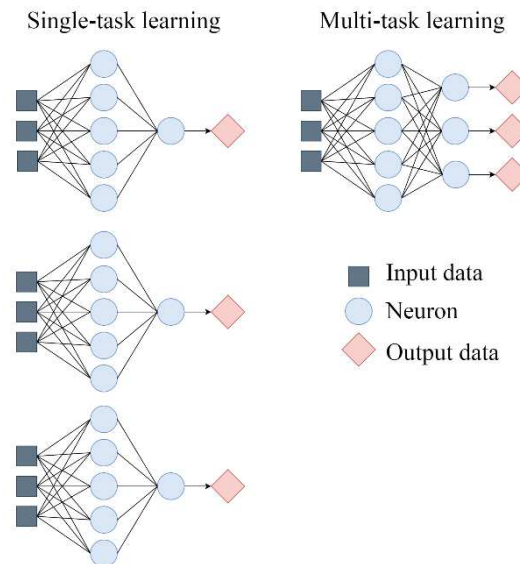
Four machine learning models, which are presented in detail later, were used: ANN, RF, XGBoost and MARS. All the models were used to estimate  $ET_o$  and missing data on maximum and minimum relative humidity (which are used to compute actual vapor pressure), solar radiation, and wind speed based on two input data combinations: temperature-based and temperature- and relative humidity-based, as shown in Fig. 3. All the models were developed considering STL. ANN and RF were developed also using MTL, since, among the models used, only these two support MTL. The missing data estimated using machine learning models were used to estimate  $ET_o$  with the FAO56-PM equation.

The following libraries for the Python programming language were used to develop the machine learning models: TensorFlow, Scikit-learn, XGBoost and py-earth. To train the models and optimize their hyperparameters, 40% of the training dataset was randomly selected to be

used as a validation subset. Thus, hyperparameters were optimized by grid-search, selecting the values of the hyperparameters that minimized the prediction errors in the validation subset.

### 2.5.1 Artificial neural network (ANN)

ANN is a machine learning model inspired by the human brain, and it is widely used in several applications, including the estimation of ETo. More details on ANN can be found in Ferreira et al. (2019). In the present study, ANN of the feed-forward multilayer perceptron (MLP) type was used. ANN models were used considering both STL and MTL. In STL, one neuron is used in the output layer, which is responsible to predict a single variable. In MTL, as multiple variables are predicted simultaneously, the number of neurons in the output layer is equal to the number of target variables. An example of using STL and MTL is presented in Fig. 4.

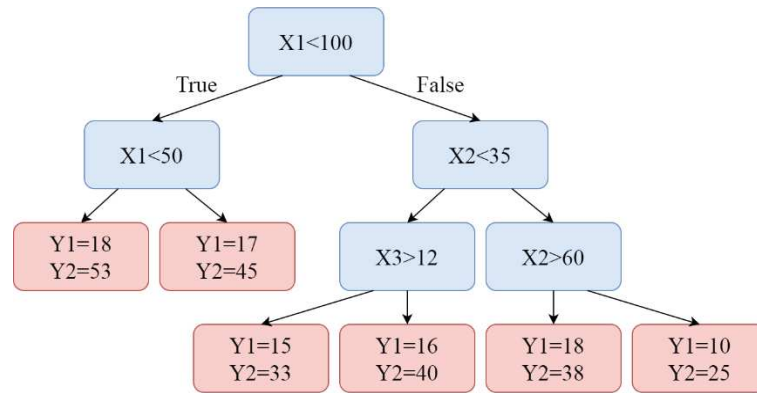


**Fig. 4.** Artificial neural network models based on single-task learning and multi-task learning.

During the training of the models, the number of hidden layers (1 and 2) and the number of neurons in each hidden layer (5, 10, 20 and 30) were optimized. The Adam training algorithm (Kingma and Ba, 2014) was employed to train the models. Hyperbolic tangent was used as activation function in hidden layers and linear function was used in the output layer. Learning rate was set to 0.001 and batch size was set to 512. The number of training epochs was defined using early stopping with maximum training epochs equal to 200 and patience equal to 20 epochs. The “restore best weights” option was also used, which restores the ANN weights that promoted the smallest errors during the last 20 epochs with no performance improvement.

### 2.5.2 Random forest (RF)

RF is a decision tree-based model. It combines several weak learners (decision trees) to produce a strong prediction (ensemble), which is obtained as the average prediction of all the decision trees in the case of regression. More details on RF can be found in Tyrallis et al. (2019). As for ANN, RF models were also developed based on STL and MTL. To allow the estimation of multiple target variables, each tree that composes a RF model has to predict all the target variables in its leaves instead of a single variable. Thus, when building the decision trees, a splitting criterion that considers all the target variables is used. It is worth mentioning that not all RF implementations support MTL. In the present study, the implementation of Scikit-learn 0.23.1 was used. To make the use of MTL in RF clearer, an example of a decision tree with two target variables is presented in Fig. 5.



**Fig. 5.** Decision tree with multi-task learning. X1, X2 and X3 - input variables 1, 2 and 3, respectively. Y1 and Y2 - target variables 1 and 2, respectively.

In hyperparameter optimization, the following hyperparameters, with their respective candidate values, were optimized: number of trees (100, 200, 400 and 600), number of features considered for splitting at each node (all features, two-thirds and one-third of the features), maximum tree depth (5, 10, 15 and 20). The minimum number of samples required to be at a leaf node was set to 10.

### 2.5.3 Extreme gradient boosting (XGBoost)

As RF, XGBoost is based on decision trees. However, it uses a different strategy to build the trees. In this algorithm, trees are created in sequence using a strategy called boosting. More information on XGBoost can be found in Chen and Guestrin (2016). In hyperparameter optimization, the following hyperparameters, with their respective candidate values, were optimized: number of trees (100, 200, 400 and 600), maximum tree depth (3, 5, 7, 9 and 11), and learning rate (0.05, 0.1, 0.2, 0.3).

### 2.5.4 Multivariate adaptive regression splines (MARS)

MARS is a model composed of base functions, which are functions fitted at different intervals of the input variables and combined to form a MARS model. Unlike most machine learning models, a MARS model is expressed as an algebraic equation, which can simplify its use for an end user. More information on MARS can be found in Cheng and Cao (2014) and Ferreira et al. (2021). In hyperparameter optimization, the following hyperparameters, with their respective candidate values, were optimized: penalty (3, 5, 10 and 20), endspace\_alpha (0.01, 0.05 and 0.1), and minspace\_alpha (0.01, 0.05 and 0.1). The order of interaction (max\_degree) was set to 3 to avoid extremely complex models.

### 2.5.5 Data normalization

Before training the models, input and output variables were standardized according to Eq. 1. This process is essential for ANN models. To avoid data leakage from the test dataset to the training dataset, the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) were computed using only data from the training subset, not including data from the validation subset and test set. Data from the validation subset and test set were standardized using  $\mu$  and  $\sigma$  obtained from the training subset.

### 2.6 Performance comparison criteria

For all the variables estimated in the present study (i.e., solar radiation, actual vapor pressure, wind speed and ETo), the models were evaluated at each weather station of the test dataset employing Nash-Sutcliffe efficiency coefficient (NSE), root mean square error (RMSE), mean bias error (MBE) and coefficient of determination ( $R^2$ ), which were calculated based on the following equations:

$$NSE = 1 - \frac{\sum (P_i - O_i)^2}{\sum (\bar{O} - O_i)^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (P_i - O_i)^2} \quad (7)$$

$$MBE = \frac{1}{n} \sum (P_i - O_i) \quad (8)$$

$$R^2 = \left[ \frac{\sum (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{(\sum (P_i - \bar{P})^2)(\sum (O_i - \bar{O})^2)}} \right]^2 \quad (9)$$

where  $P_i$  is the predicted value,  $O_i$  is the observed value,  $\bar{P}$  is the mean of the predicted values,  $\bar{O}$  is the mean of the observed values, and  $n$  is the number of data pairs.

NSE indicates the magnitude of the mean square error ( $RMSE^2$ ) in relation to the variance of the observed data. RMSE is a general error indicator that gives more weight to large errors. MBE is used to represent the general tendency of a model to underestimate or overestimate observed values.  $R^2$  indicates the association between observed and predicted values. NSE and

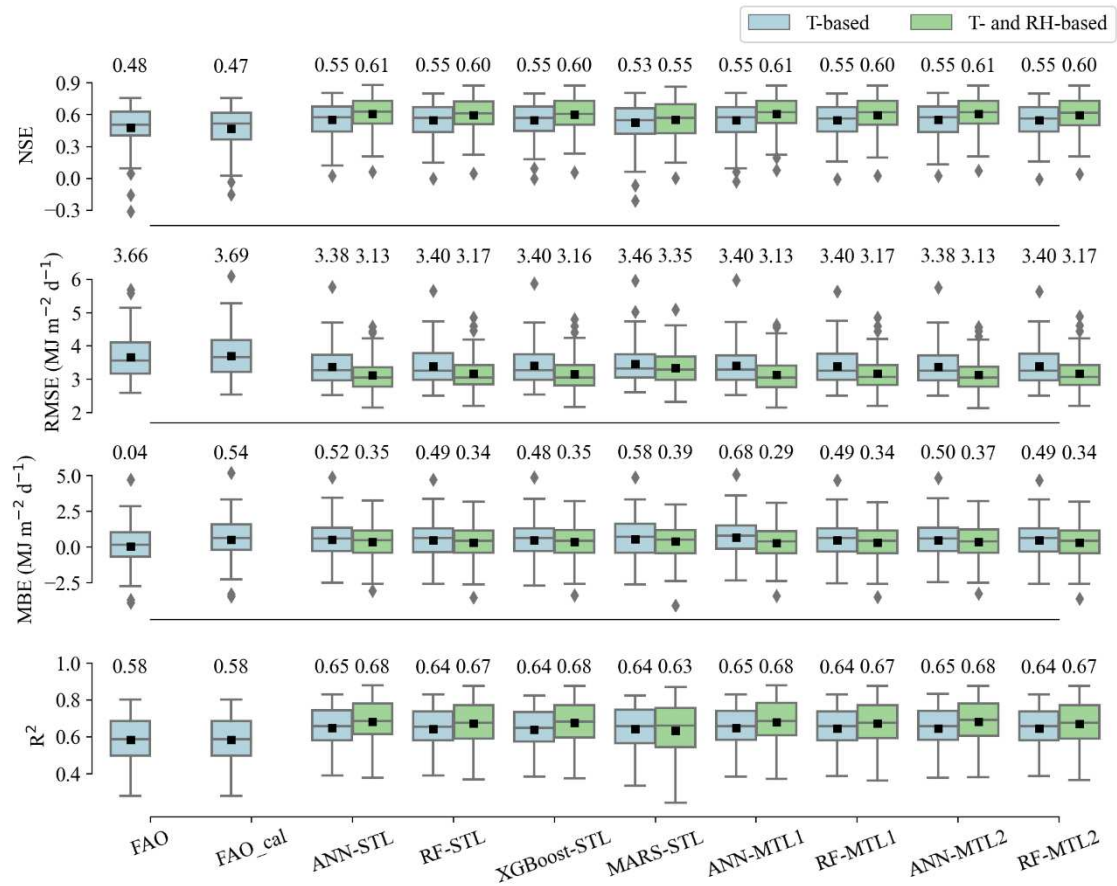
$R^2$  should be as close to one as possible and RMSE and MBE should be as close to zero as possible.

### 3 Results and discussion

#### 3.1 General scenario

##### 3.1.1 Solar radiation estimation

The performance of the machine learning models and the original and calibrated FAO-56 approach to estimate solar radiation is presented in Fig. 6. Overall, all the machine learning models performed similarly, with slightly poorer results for the MARS model. In general, the use of MTL showed results very close to those obtained using STL. The use of relative humidity in addition to air temperature promoted performance gains, increasing NSE and  $R^2$  and decreasing RMSE and MBE (absolute values). For the temperature-based models, the best performing options were the ANN-STL and ANN-MTL2 (NSE: 0.55, RMSE: 3.38 and  $R^2$ : 0.65). For the temperature- and relative humidity-based models, the ANN-STL, ANN-MTL1 ANN-MTL2 exhibited the best performances, both with NSE: 0.61, RMSE: 3.13, and  $R^2$ : 0.68. When compared with the original and calibrated FAO approach, all the machine learning models performed better.





**Fig. 6.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for different models used to estimate solar radiation. T - temperature; RH - relative humidity; FAO\_cal - calibrated FAO; ANN - Artificial neural network; RF - Random forest; XGBoost - Extreme gradient boosting; MARS - Multivariate adaptive regression splines; STL - single-task learning; MTL - multi-task learning. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

Regarding the calibration of the FAO method, in contrast to our initial expectations, the original model (NSE: 0.48, RMSE: 3.66, MBE: 0.04 and  $R^2$ : 0.58) performed slightly better than its calibrated version (NSE: 0.47, RMSE: 3.69, MBE: 0.54 and  $R^2$ : 0.58). The change in  $k_{RS}$  value was small, from 0.16 to 0.1643. The slight performance loss observed is probably due to some random differences between the training and test datasets. Thus, even promoting better results in the training dataset, the calibrated  $k_{RS}$  performed a little worse in the test dataset.

The performance improvement observed when adding relative humidity as input data was also observed in other studies (He et al., 2020; Quej et al., 2016; Valiantzas, 2018b). Valiantzas (2018b) demonstrated empirically that mean relative humidity has an inverse relation with solar radiation. Lindauer et al. (2017) provided evidence that relative humidity has a certain relation with cloud cover and atmospheric transmissivity, which are directly related to solar radiation at the land surface.

### 3.1.2 Actual vapor pressure estimation

The machine learning models showed similar performances to estimate actual vapor pressure (computed from maximum and minimum relative humidity estimated by the machine learning models), with only slightly poorer results for the MARS model (Fig. 7). When using MTL, the results were very close to those obtained using STL. The best performing model was the ANN-MTL2 (NSE: 0.40, RMSE: 0.23, MBE: -0.02 and  $R^2$ : 0.71). However, the other models had similar results. Although the FAO approach did not achieve the same performance obtained by the machine learning models, it was benefited from the calibration, increasing NSE from -1.33 to -0.12, decreasing RMSE from 0.42 to 0.32, and decreasing MBE from 0.32 to 0.02. It is observed that there was an actual vapor pressure overestimation (MBE: 0.32) before the calibration, which was reduced by subtracting 2.21°C from minimum air temperature during actual vapor pressure computation (Eq. 3). However, even after the calibration, a negative mean NSE value was obtained, which indicates that the overall performance was poor.



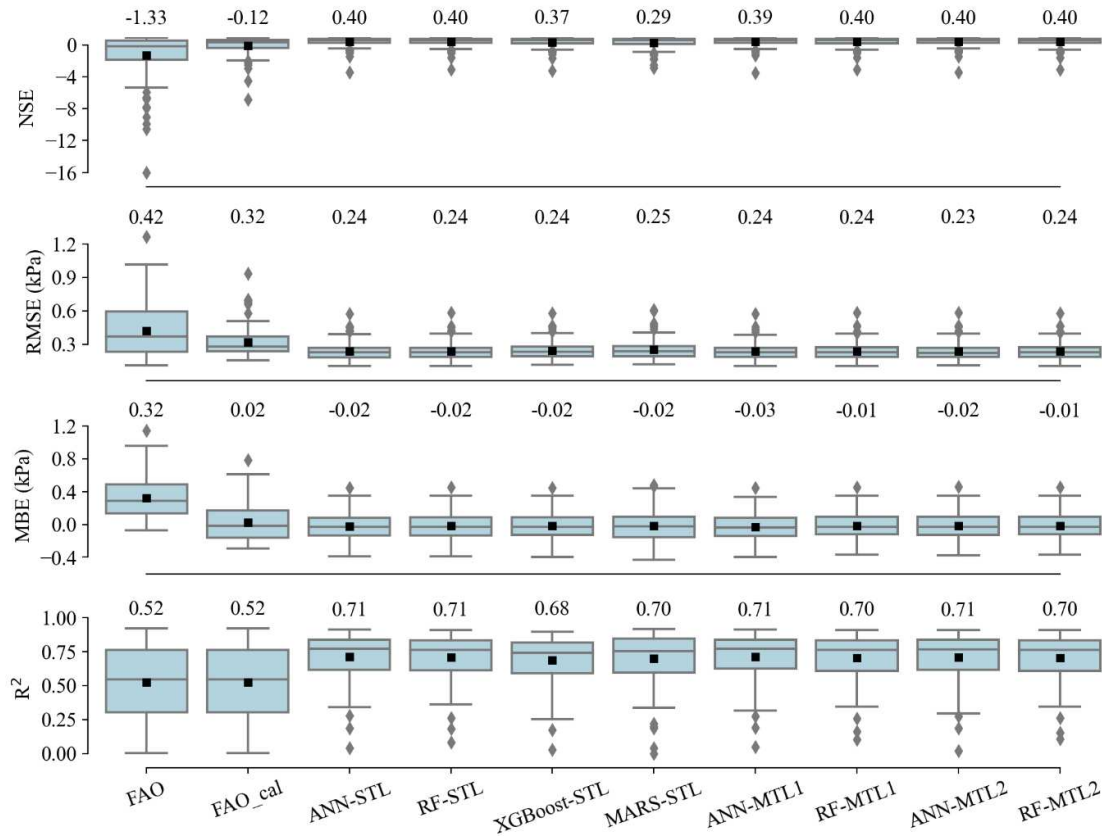


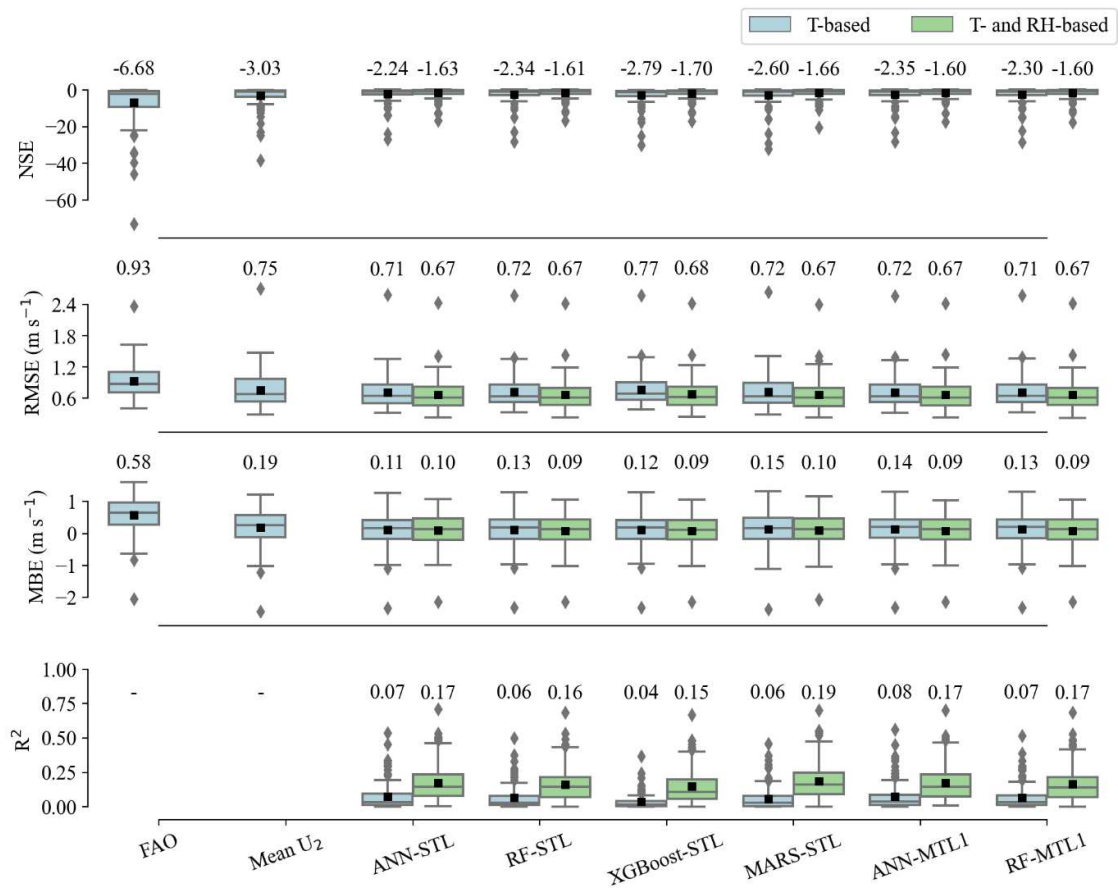
Fig.

7. Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for different models used to estimate actual vapor pressure. FAO\_cal - calibrated FAO; ANN - Artificial neural network; RF - Random forest; XGBoost - Extreme gradient boosting; MARS - Multivariate adaptive regression splines; STL - single-task learning; MTL - multi-task learning. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

When estimating actual vapor pressure from air temperature, minimum temperature is considered an approximation of dew point temperature. However, it should be adjusted according to the local climatic conditions by subtracting a constant value (Allen et al., 1998; Paredes and Pereira, 2019). According to our study, overall, for Brazil, subtracting 2.21°C improved the estimation. However, it is important to note that it is a general recommendation, since this value varies according to the local climatic conditions, being lower for humid regions and higher for dryer regions. In addition, in humid regions, minimum temperature can be higher than dew point temperature (Paredes and Pereira, 2019).

### 3.1.3 Wind speed estimation

The machine learning models performed similarly when predicting wind speed, with slightly poorer results for the XGBoost model (Fig. 8). As observed for the previous variables, in general, the use of MTL promoted results similar to those found using STL. The use of relative humidity in addition to air temperature promoted performance gains. However, all the models were not able to predict wind speed with adequate accuracy, obtaining negative mean NSE values. For the temperature-based models, the best performing option was the ANN-STL (NSE: -2.24, RMSE: 0.71, MBE: 0.11 and  $R^2$ : 0.07). For the temperature- and relative humidity-based models, the best performing options were the ANN-MTL1 and RF-MTL1 (NSE: -1.60, RMSE: 0.67, MBE: 0.09 and  $R^2$ : 0.17). The use of the average wind speed obtained at the training stations ( $1.61 \text{ m s}^{-1}$ ) promoted better results (NSE: -3.03, RMSE: 0.75 and MBE: 0.19) than the world average wind speed value ( $2.00 \text{ m s}^{-1}$ ) proposed in the FAO approach (NSE: -6.68, RMSE: 0.93 and MBE: 0.58). Paredes et al. (2018a) and Popova et al. (2006) reported better results when using local/regional average wind speed instead of the default value proposed in the FAO approach in estimating ETo.



**Fig. 8.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for different models used to estimate wind speed. T - temperature; RH - relative humidity; ANN - Artificial neural network; RF - Random forest; XGBoost - Extreme gradient boosting; MARS -

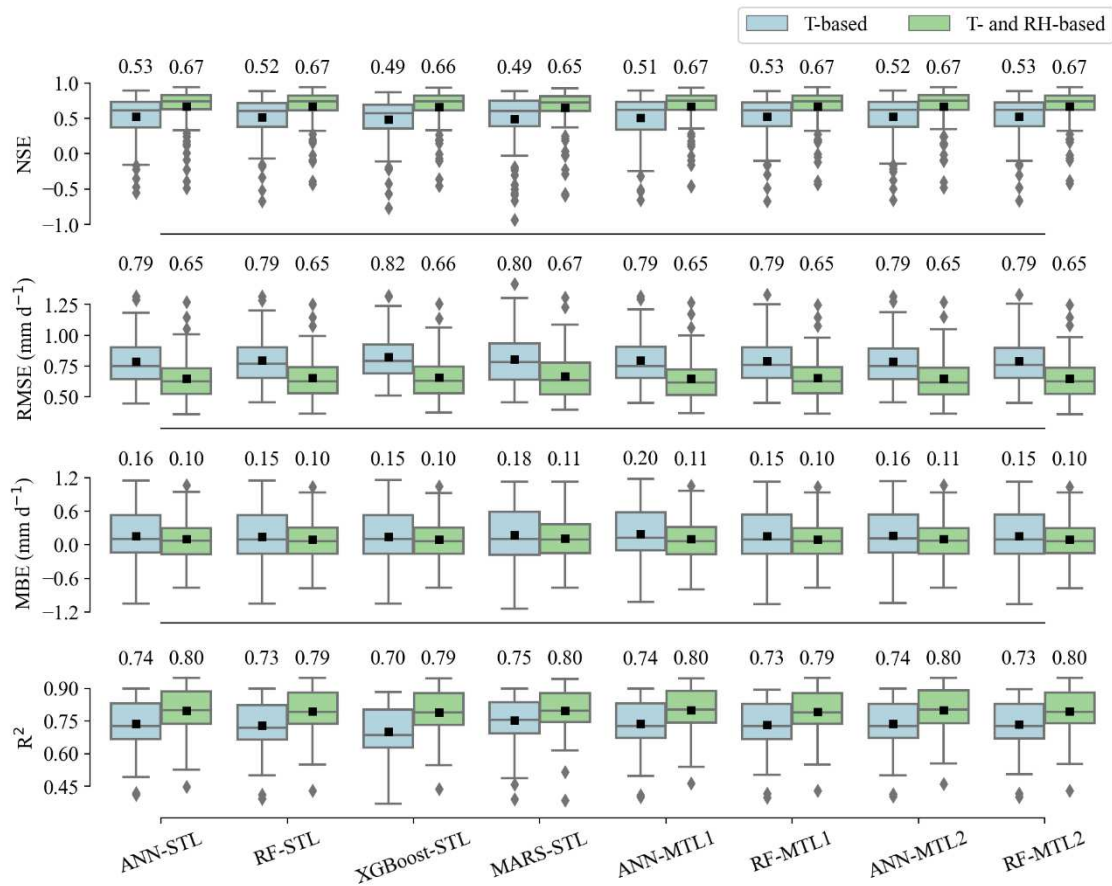
Multivariate adaptive regression splines; STL - single-task learning; MTL - multi-task learning.  
 \*As FAO and Mean  $U_2$  models provide a constant value as prediction for wind speed,  $R^2$  values cannot be computed. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

The poor performance in estimating wind speed is associated with the low correlation between wind speed and the input data considered for developing the models (i.e., air temperature, relative humidity and extraterrestrial solar radiation). However, even with the aforementioned low correlation, the machine learning models were able to capture some patterns between input and output data and produce lower errors than the simple use of a constant value (FAO and average wind speed approaches) (Fig. 8).

### **3.1.4 Reference evapotranspiration estimation**

#### **3.1.4.1 Machine learning models**

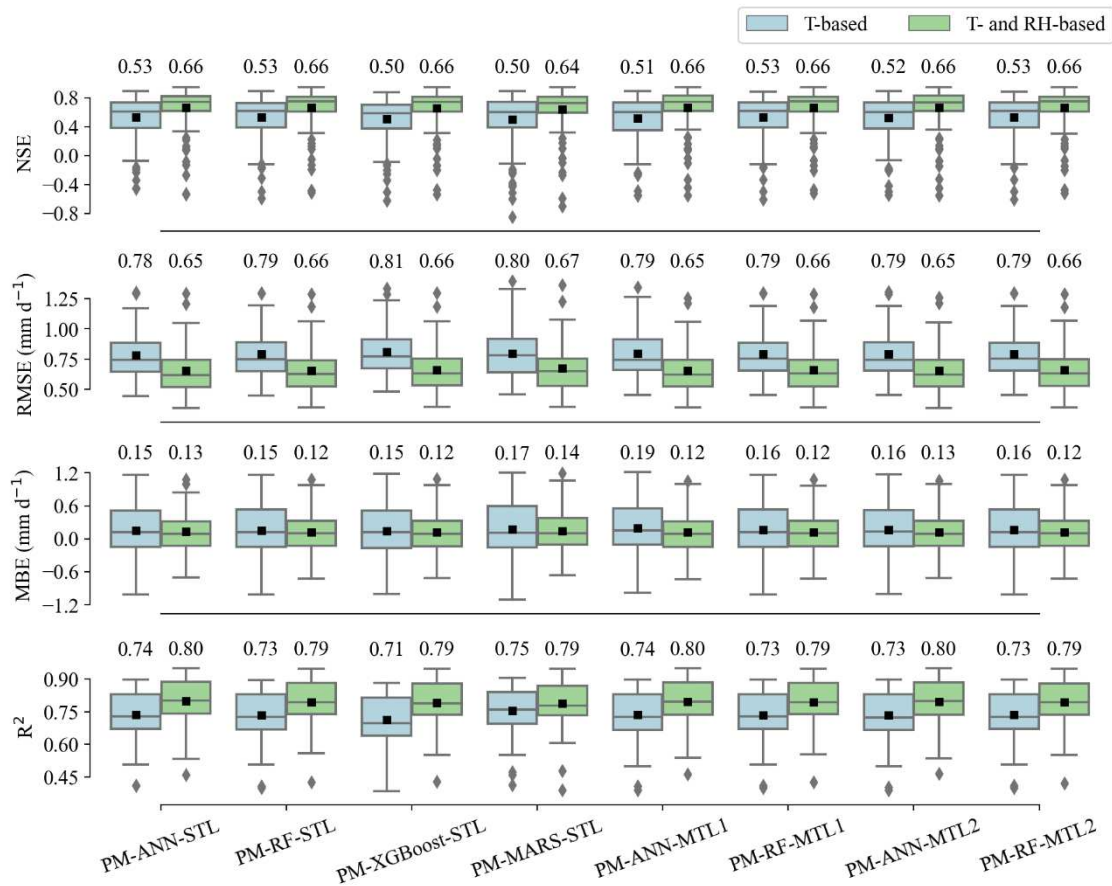
The performance of the machine learning models used to estimate ETo directly from air temperature and relative humidity or temperature only is presented in Fig. 9. The models performed similarly, with a slight advantage for the ANN and RF models. As previously observed, in general, the use of MTL did not provide strong performance changes, showing similar results to those obtained with STL. Adding relative humidity as input resulted in performance gains. For the temperature-based models, the best performing models were the ANN-STL, RF-MTL1 and RF-MTL2 (NSE: 0.53 and RMSE: 0.79). For the temperature- and relative humidity-based models, the ANN and RF developed using STL, MTL1 and MTL2 showed the best performances (NSE: 0.67 and RMSE: 0.65).



**Fig. 9.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for the machine learning models used to estimate ET<sub>o</sub>. T - temperature; RH - relative humidity; ANN - Artificial neural network; RF - Random forest; XGBoost - Extreme gradient boosting; MARS - Multivariate adaptive regression splines; STL - single-task learning; MTL - multi-task learning. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

### 3.1.4.2 FAO56-PM equation combined with machine learning models

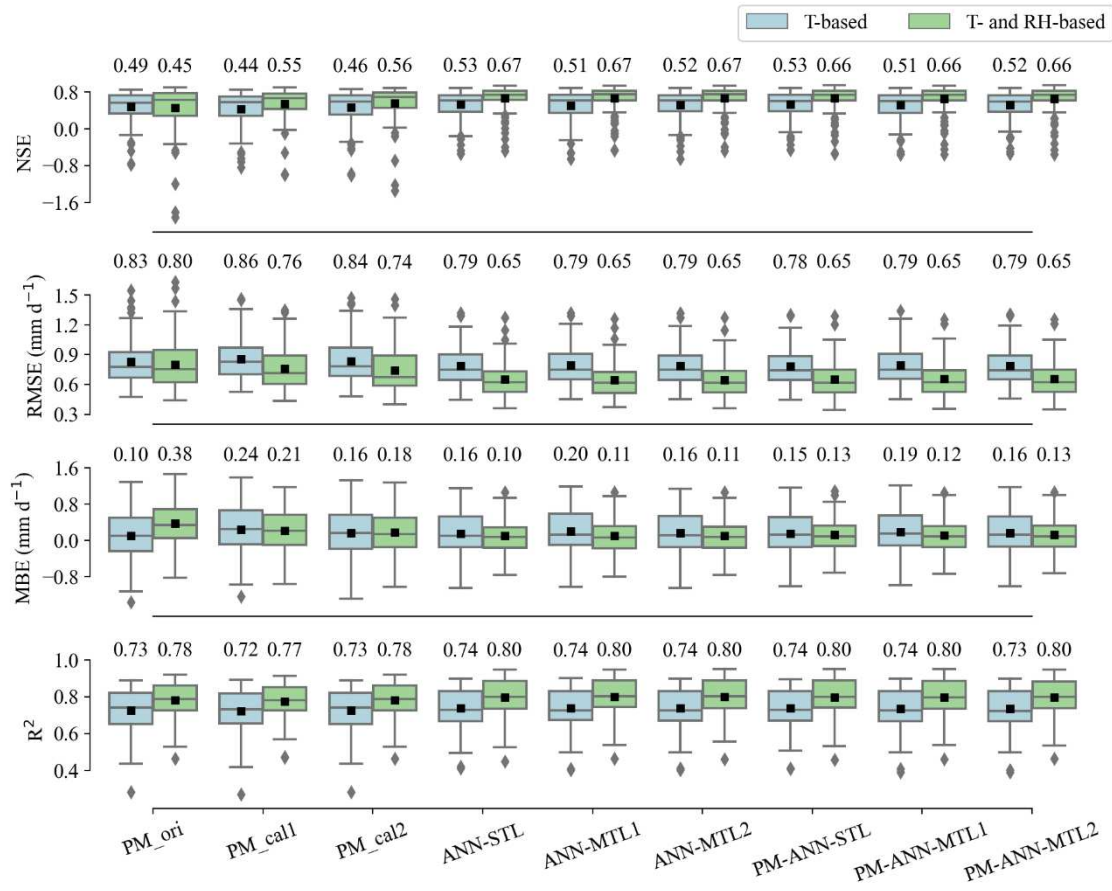
The performance of the FAO56-PM equation combined with machine learning models (used to estimate missing data) is presented in Fig. 10. All the models had similar performance, with a slight advantage for those that used ANN and RF. The models developed employing MTL and STL performed similarly. As previously observed, there were also performance gains by adding relative humidity as input. For the temperature-based models, the best performing option was the PM-ANN-STL (FAO56-PM combined with ANN-STL) (NSE: 0.53, RMSE: 0.78, MBE: 0.15 and  $R^2$ : 0.74). For the temperature- and relative humidity-based models, the FAO56-PM combined with the ANN models developed with STL, MTL1 and MTL2 performed the best (NSE: 0.66, RMSE: 0.65 and  $R^2$ : 0.80).



**Fig. 10.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for ETo estimated using the FAO56-PM equation with missing data obtained employing machine learning models. T - temperature; RH - relative humidity; PM - FAO56-PM equation; ANN - Artificial neural network; RF - Random forest; XGBoost - Extreme gradient boosting; MARS - Multivariate adaptive regression splines; STL - single-task learning; MTL - multi-task learning. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

### 3.1.5 Overall evaluation

As presented before, the ANN and RF models used individually or combined with the FAO56-PM equation provided the best performances, with a slight advantage for ANN. Thus, to better compare the different approaches used to estimate ETo, the performance of the FAO56-PM equation with missing data estimated based on traditional methodologies and ANN models developed considering single-task and multi-task learning, individually used and combined with FAO56-PM equation, is presented in Fig. 11.



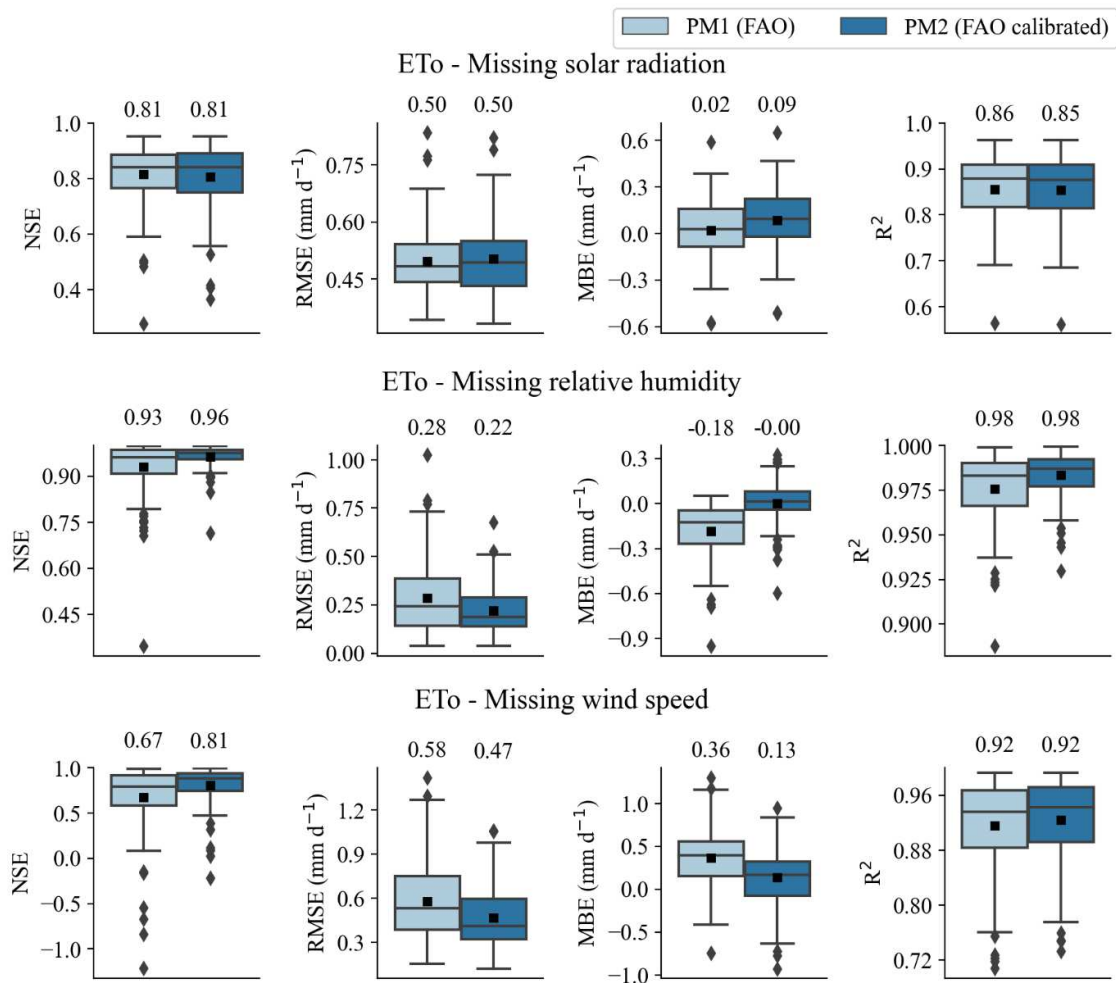
**Fig. 11.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for ETo estimated with the best model from each approach assessed (FAO56-PM equation, machine learning, and FAO56-PM equation combined with machine learning). T - temperature; RH - relative humidity; PM - FAO56-PM equation; PM\_ori - PM with missing data estimated using original FAO methodologies; PM\_cal1 - PM with missing data estimated using calibrated FAO methodologies; PM\_cal2 - PM with missing data estimated using original FAO methodologies and calibrated using linear regression; ANN - Artificial neural network; STL - single-task learning; MTL - multi-task learning. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

When using only measured data on temperature, the FAO56-PM equation with missing data estimated based on FAO methodologies (PM\_ori, PM\_cal1 and PM\_cal2 in Fig. 11) performed better before both evaluated calibrations. In contrast, it performed better after the calibrations when using temperature and relative humidity as input data. The performance loss observed for the temperature-based equation is not expected, since calibrations generally promote performance gains. In addition, regarding PM\_cal1, when evaluating the estimation of missing data after the calibrations, although there was a slight decrease in the performance of



the estimation of solar radiation, the estimations of actual vapor pressure and wind speed were improved. Thus, both calibration types evaluated (PM\_cal1 and PM\_cal2) are expected to improve the estimation of ETo.

To better investigate the cause of the performance loss in the temperature-based FAO56-PM equation after calibrating the methodologies used to estimate missing data (PM\_cal1) (Fig. 11), the performance of the FAO56-PM equation using each one of the variables estimated (i.e., solar radiation, relative humidity/actual vapor pressure, and wind speed) as the sole missing data was assessed (Fig. 12). After the calibrations, better ETo estimates were obtained when considering actual vapor pressure and wind speed as missing data. When considering solar radiation as missing data, after the calibration, a similar performance was observed, with only a small increase in MBE (0.02 to 0.09) and a small decrease in  $R^2$  (0.86 to 0.85). Based on these results, the calibrations are expected to improve the performance of the FAO56-PM equation when using only measured data on temperature, since only the performance of the estimation of solar radiation had a minimal decrease. However, a contrary behavior was observed (Fig. 11). It is probably justified by the characteristics of the errors in the missing data estimated. Probably, although the errors obtained with the non-calibrated FAO methodologies to estimate missing data are higher, they were partially canceled out when considering the use of solar radiation, actual vapor pressure and wind speed estimates simultaneously, which justify the better performance of the FAO56-PM equation with the non-calibrated FAO methodologies (PM\_ori).



**Fig. 12.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for ETo estimated using the FAO56-Penman-Monteith equation with missing data on solar radiation, relative humidity, and wind speed, which were estimated using the original and calibrated FAO methodologies. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

Calibrating the FAO56-PM equation with missing data estimated using the original FAO methodologies through linear regression (PM\_cal2) provided slightly better results in relation to PM\_cal1 in both temperature-based and temperature- and relative-humidity-based settings (Fig. 11). Although calibrating individual equations to estimate missing data can have a more in-depth effect, the calibration based on linear regression has the advantage of adjusting the model with a focus on the estimation of ETo itself. As previously mentioned, in the temperature-based setting, even promoting better results than PM\_cal1, the PM\_cal2 model also performed slightly worse than the PM\_ori model. It probably occurred due to some random difference between data used to calibrate and test the equation.



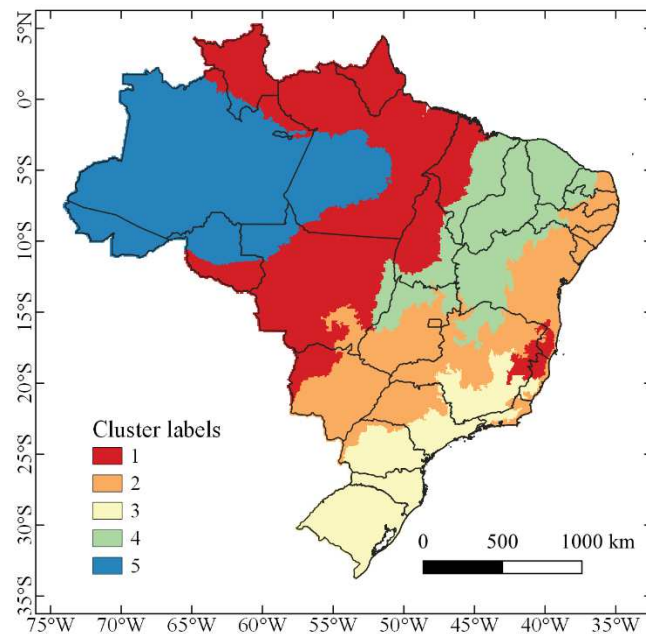
Compared to the FAO56-PM equation with missing data estimated using traditional methodologies (i.e., PM\_ori, PM\_cal1 and PM\_cal2 in Fig. 11), the machine learning models and their combinations with the FAO56-PM equation performed better to estimate ETo, mainly the models based on temperature and relative humidity (Fig. 11). For the temperature-based models, when comparing the non-calibrated PM model (PM\_ori), which performed better than its calibrated versions (PM\_cal1 and PM\_cal2), with the PM-ANN-STL model, which was the best machine learning model, NSE increased from 0.49 to 0.53 (8.2%), RMSE decreased from 0.83 to 0.78 (6.0%), and  $R^2$  increased from 0.73 to 0.74 (1.4%). For the temperature- and relative humidity-based models, when comparing the PM\_cal2 model, which performed better than its other versions, with the ANN-STL, ANN-MTL1 and ANN-MTL2 models or RF-STL, RF-MTL1 and RF-MTL2 (Fig. 9), which were the best machine learning models, NSE increased from 0.56 to 0.67 (19.6%) and RMSE decreased from 0.74 to 0.65 (12.2%).

Overall, the use of relative humidity in addition to temperature improved the performance of the models studied. However, for the non-calibrated PM model (PM\_ori), although adding relative humidity decreased RMSE from 0.83 to 0.80 (3.6%), NSE reduced from 0.49 to 0.45 (8.2%) (Fig. 11). This unexpected behavior observed for the PM\_ori model is probably due to the actual vapor pressure overestimation, which occurred when using only temperature (Fig. 7 (MBE: 0.32)), which is, at a certain point, canceled out by the overestimated wind speed (Fig. 8 (MBE: 0.58)). It occurs since the actual vapor pressure overestimation led to ETo underestimation and the wind speed overestimation led to ETo overestimation, canceling out part of the errors. Therefore, when relative humidity data were added to the PM\_ori model, the mentioned condition was no more achieved, which can contribute to increase the errors in the estimation of ETo at some weather stations. However, when a better wind speed value was considered (i.e., PM\_cal1), adding relative humidity was translated into performance gains, increasing NSE from 0.44 to 0.55 (25.0%). Similarly, for the model calibrated using linear regression (PM\_cal2), NSE was increased from 0.46 to 0.56 (21.7%) when adding relative humidity. By comparing the best temperature-based and temperature- and relative humidity-based models to estimate ETo (i.e., PM-ANN-STL and ANN-STL/ANN-MTL1/ANN-MTL2 (or RF developed with STL, MTL1 and MTL2), respectively), NSE increased from 0.53 to 0.67 (26.4%) and RMSE decreased from 0.78 to 0.65 (16.7%).

### 3.2 Clustering scenario

As mentioned before, ANN and RF provided the best performances in estimating missing data and ETo, with a slight advantage for ANN. Thus, to better explore the potential of ANN models and the FAO56-PM equation, we also developed/calibrated them after dividing Brazil

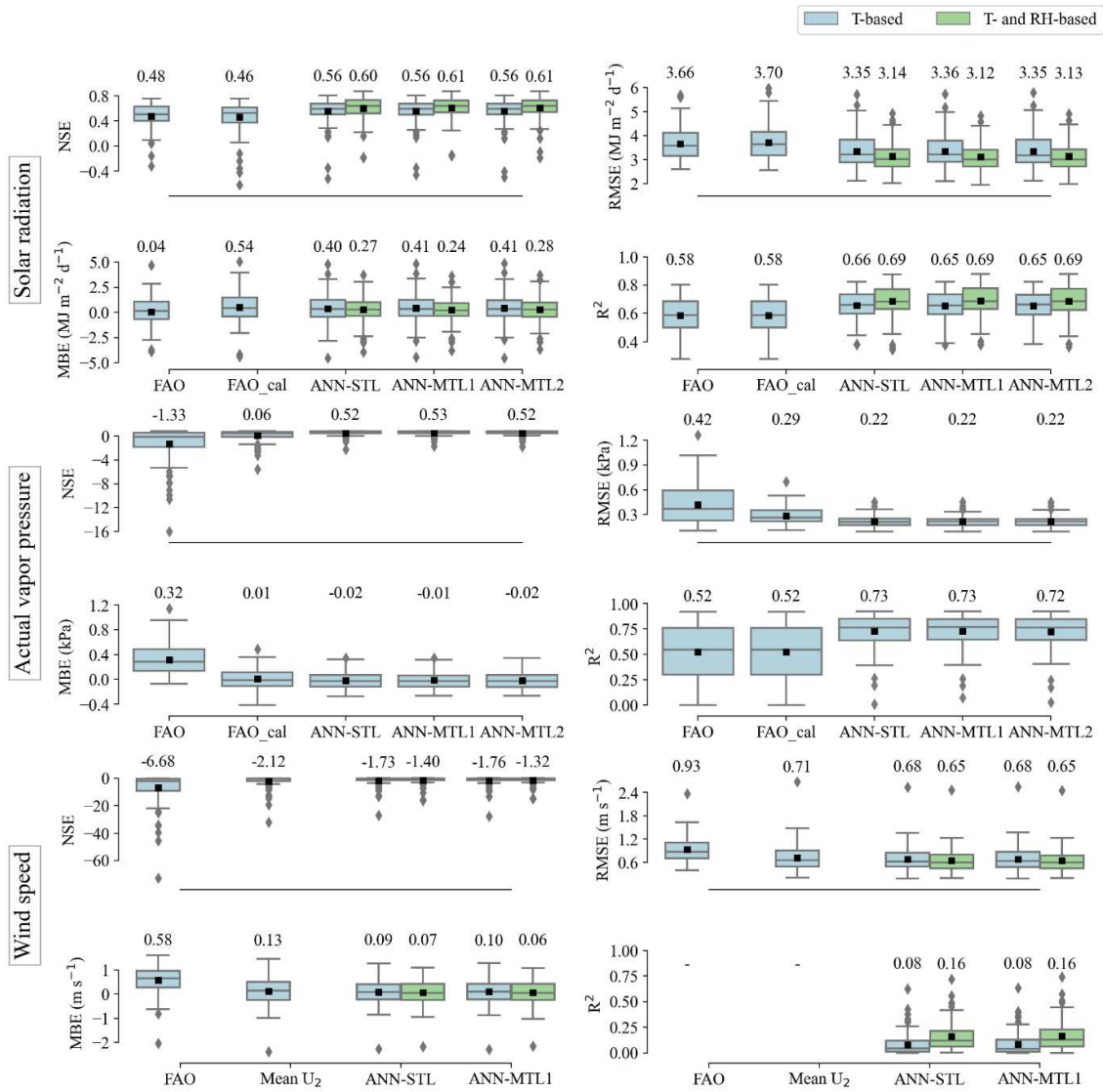
into regions with similar climatic conditions (clustering). The clusters obtained using K-means and data from WorldClim are presented in Fig. 13.



**Fig. 13.** Regions (clusters) with similar climatic conditions defined using K-means.

### 3.2.1 Missing data and reference evapotranspiration

The performance of the models developed for each cluster to estimate solar radiation, actual vapor pressure and wind speed is presented in Fig. 14. As seen in the general scenario, the ANN models outperformed the FAO methodologies used to estimate missing data, in their original and calibrated forms, and STL and MTL provided similar performances.

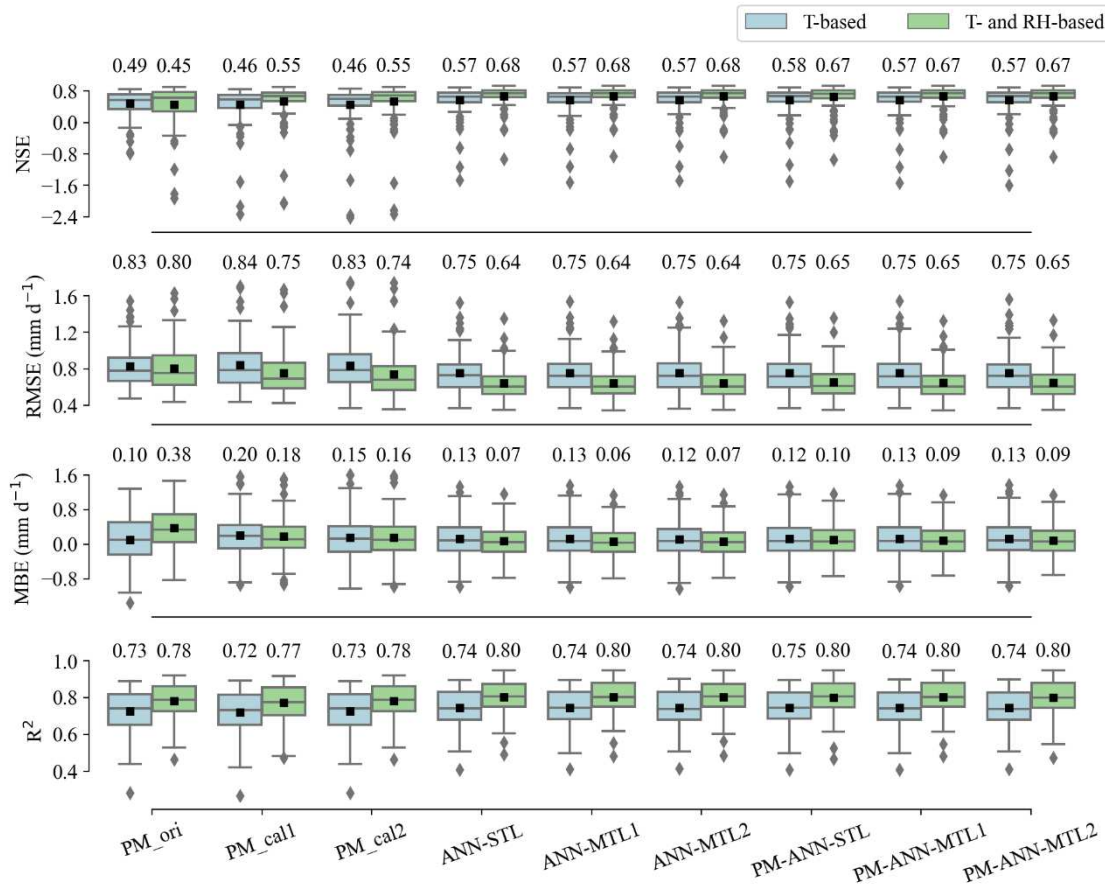


**Fig. 14.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for solar radiation, actual vapor pressure and wind speed estimated using different models developed for groups of weather stations with similar climatic conditions (clustering). T - temperature; RH - relative humidity; FAO\_cal - calibrated FAO; ANN - Artificial neural network; STL - single-task learning; MTL - multi-task learning. \*As FAO and Mean  $U_2$  models provide a constant value as prediction for wind speed,  $R^2$  values cannot be computed. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

After the clustering, similar results were found for solar radiation (Figs. 6 and 14). For actual vapor pressure (Figs. 7 and 14), there were considerable performance improvements, increasing the best NSE value from 0.40 to 0.53. However, in the case of the calibrated FAO methodology, even increasing NSE from -0.12 to 0.06, the performance was still poor, since

NSE was very close to zero. Thus, better approaches to calibrate the FAO methodology used to estimate actual vapor pressure for Brazil are necessary. For wind speed (Figs. 8 and 14), although the clustering increased NSE and decreased RMSE, all models continued to have negative NSE values, which indicates their poor performance. Thus, different approaches should be explored in the future to produce more reliable wind speed estimates.

By evaluating ETo estimated after the clustering (Figure 15), as observed in the general scenario, the ANN models used individually and in combination with the FAO56-PM equation outperformed the FAO56-PM equation with missing data estimated based on traditional methodologies (original and calibrated). Using STL and MTL provided similar results, as found in the general scenario.



**Fig. 15.** Boxplots and mean values of NSE, RMSE, MBE and  $R^2$  (for the test dataset) for ETo estimated using different models developed for groups of weather stations with similar climatic conditions (clustering). T - temperature; RH - relative humidity; PM - FAO56-PM equation; PM\_ori - PM with missing data estimated using original FAO methodologies; PM\_cal1 - PM with missing data estimated using calibrated FAO methodologies; PM\_cal2 - PM with missing data estimated using original FAO methodologies and calibrated using linear regression; ANN - Artificial neural network; STL - single-task learning; MTL - multi-task

learning. Means are numerically represented by the values at the top of each boxplot and by the black squares. Outliers are represented by the gray lozenges.

Regarding the calibration of the FAO56-PM equation, as observed in the general scenario, in the temperature-based setting, the original equation performed slightly better than its calibrated version. However, in the temperature- and relative humidity-based setting, both calibrations strategies promoted performance gains. Although PM\_cal1 and PM\_cal2 performed similarly, PM\_cal2 showed slightly lower RMSE values.

The optimal values of the calibration parameters used with the FAO56-PM equation are provided in Table 1. Cluster 5 showed the lowest values of  $a_T$ ,  $k_{RS}$  and  $U_2$ . This cluster is located mainly in northern Brazil, where Amazon rainforest predominates. This region is characterized by Af and Am climates (Fig. 1) and high relative humidity and rainfall, in addition to low wind speed. Thus, minimum temperature tends to better approximate dew point temperature, which justify the lowest  $a_T$  value. The high rainfall levels, which are related to high cloud cover, reduce the proportion of solar radiation at surface in relation to extraterrestrial solar radiation, which corroborates the lower  $k_{RS}$  value. On the other hand, cluster 4 showed the highest values of  $a_T$  and  $k_{RS}$ . This cluster is located in a dryer region, with Aw, As and BSh climates (Fig. 1). Thus, under these conditions,  $a_T$  is expected to assume a higher value to make minimum temperature closer to dew point temperature, as found in this study. In addition, due to the lower rainfall levels (lower cloud cover), the proportion of solar radiation at surface in relation to extraterrestrial solar radiation is expected to be higher, which increases the  $k_{RS}$  value. Clusters 1, 2 and 3 exhibited similar values of  $a_T$ ,  $k_{RS}$  and  $U_2$ .

**Table 1.** Optimal values of the calibration parameters used with the FAO56-PM equation for each cluster and for the general calibration.

Input type	Parameter	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	General
-	$a_T$	1.91	2.31	1.22	4.21	0.94	2.21
	$k_{RS}$	0.1574	0.1655	0.163	0.174	0.153	0.1643
	$U_2$	1.14	1.65	1.84	1.79	0.70	1.61
T	a	0.6356	0.0739	-0.217	0.9036	-0.327	-0.0671
	b	0.7897	1.0134	1.0588	0.934	0.9407	1.0292
T and RH	a	0.2788	-0.1314	-0.2592	-0.0671	-0.3418	-0.1988
	b	0.8231	0.9963	1.0344	1.0326	0.9143	1.0009

T - temperature; RH - relative humidity;  $a_T$  is the minimum temperature correction factor used in actual vapor pressure estimation;  $k_{RS}$  is an empirical radiation adjustment coefficient;  $U_2$  is the mean wind speed at a 2 m height ( $m\ s^{-1}$ ); a and b are the calibrations parameters used in the calibration based on simple linear regression.

### 3.3 Comparison of the general and clustering scenarios

To compare the performance of the models developed to estimate ETo for Brazil as a whole and for regions with similar climatic conditions (clusters), mean values of NSE, RMSE, MBE and  $R^2$  over test stations are presented in Table 2. The temperature-based models performed better after the clustering. In general, after the clustering, mean NSE over the models assessed increased from 0.50 to 0.54 (8.0%), RMSE reduced from 0.81 to 0.78 (3.7%), MBE reduced from 0.17 to 0.14 (17.6%) and  $R^2$  increased from 0.73 to 0.74 (1.4%). The best performance was obtained by the PM-ANN-STL model, with NSE equal to 0.58 and RMSE equal to 0.75. In contrast, the temperature- and relative humidity-based models had similar performances in the general and clustering scenarios. Models developed to estimate ETo based on limited input data are influenced by the local climatic conditions. Therefore, when developing models for regions with more homogeneous characteristics, they tend to better capture the relations between input data and ETo, attaining better performance than models developed for heterogeneous regions. Ferreira et al. (2019) also found better performances when using a clustering strategy in Brazil. These authors reported higher performance gains for temperature-based models in relation to temperature- and relative humidity-based models, as observed in the present study. In contrast, Althoff et al. (2019) reported only small gains by clustering weather stations. However, it may have occurred due to the smaller study area (lower climatic variability) considered in the mentioned study in relation to the one considered in the present study.

**Table 2.** Mean values of different error metrics (NSE, RMSE, MBE and  $R^2$ ) for the models developed to estimate ETo for Brazil as a whole (General) and for groups of weather stations (Clustering). Highlighted values indicate the best error metrics values for each cluster.

Input type	Model	General				Clustering			
		NSE	RMSE	MBE	$R^2$	NSE	RMSE	MBE	$R^2$
T	PM_ori	0.49	0.83	<b>0.10</b>	0.73	0.49	0.83	<b>0.10</b>	0.73
	PM_cal1	0.44	0.86	0.24	0.72	0.46	0.84	0.20	0.72
	PM_cal2	0.46	0.84	0.16	0.73	0.46	0.83	0.15	0.73
	ANN-STL	<b>0.53</b>	0.79	0.16	<b>0.74</b>	0.57	<b>0.75</b>	0.13	0.74



	ANN-MTL1	0.51	0.79	0.20	<b>0.74</b>	0.57	<b>0.75</b>	0.13	0.74
	ANN-MTL2	0.52	0.79	0.16	<b>0.74</b>	0.57	<b>0.75</b>	0.12	0.74
	PM-ANN-STL	<b>0.53</b>	<b>0.78</b>	0.15	<b>0.74</b>	<b>0.58</b>	<b>0.75</b>	0.12	<b>0.75</b>
	PM-ANN-MTL1	0.51	0.79	0.19	<b>0.74</b>	0.57	<b>0.75</b>	0.13	0.74
	PM-ANN-MTL2	0.52	0.79	0.16	0.73	0.57	<b>0.75</b>	0.13	0.74
	Mean	0.50	0.81	0.17	0.73	0.54	0.78	0.14	0.74
T and RH	PM_ori	0.45	0.80	0.38	0.78	0.45	0.80	0.38	0.78
	PM_cal1	0.55	0.76	0.21	0.77	0.55	0.75	0.18	0.77
	PM_cal2	0.56	0.74	0.18	0.78	0.55	0.74	0.16	0.78
	ANN-STL	<b>0.67</b>	<b>0.65</b>	<b>0.10</b>	<b>0.80</b>	<b>0.68</b>	<b>0.64</b>	0.07	<b>0.80</b>
	ANN-MTL1	<b>0.67</b>	<b>0.65</b>	0.11	<b>0.80</b>	<b>0.68</b>	<b>0.64</b>	<b>0.06</b>	<b>0.80</b>
	ANN-MTL2	<b>0.67</b>	<b>0.65</b>	0.11	<b>0.80</b>	<b>0.68</b>	<b>0.64</b>	0.07	<b>0.80</b>
	PM-ANN-STL	0.66	<b>0.65</b>	0.13	<b>0.80</b>	0.67	0.65	0.10	<b>0.80</b>
	PM-ANN-MTL1	0.66	<b>0.65</b>	0.12	<b>0.80</b>	0.67	0.65	0.09	<b>0.80</b>
	PM-ANN-MTL2	0.66	<b>0.65</b>	0.13	<b>0.80</b>	0.67	0.65	0.09	<b>0.80</b>
	Mean	0.62	0.69	0.16	0.79	0.62	0.69	0.13	0.79

PM - FAO56-PM equation; PM\_ori - PM with missing data estimated using original FAO methodologies; PM\_cal1 - PM with missing data estimated using calibrated FAO methodologies; PM\_cal2 - PM with missing data estimated using original FAO methodologies and calibrated using linear regression; ANN - Artificial neural network; STL - single-task learning; MTL - multi-task learning; NSE - Nash-Sutcliffe efficiency coefficient; RMSE - root mean square error; MBE - mean bias error;  $R^2$  coefficient of determination.

As the temperature- and relative humidity-based models used to estimate ETo were little affected by the clustering, it can indicate that the regions (clusters) created were not efficient to overcome the missing of solar radiation and wind speed data. It is also supported by the results presented in Figure 6, 7, 8 and 14, in which it can be noted that, in contrast to actual vapor pressure, solar radiation estimation was generally not improved after the clustering and, although wind speed estimation had some improvement, it continues to obtain poor predictions after the clustering. On the other hand, actual vapor pressure estimation had a considerable performance increase.

### 3.4 Overall evaluation

In both general and clustering scenarios, the use of MTL did not provide strong performance changes in estimating ETo, showing similar results to those found using STL (Figs. 9, 10 and 15). The same behavior was observed in estimating the missing data considered



(Figs. 6, 7, 8 and 14). However, it is worth mentioning that when in addition to the estimation of ETo the estimations of solar radiation, actual vapor pressure (or relative humidity) and wind speed are of interest, the use of MTL can reduce the time and computational resources required to develop the machine learning models, since only a single model is required to estimate all these variables.

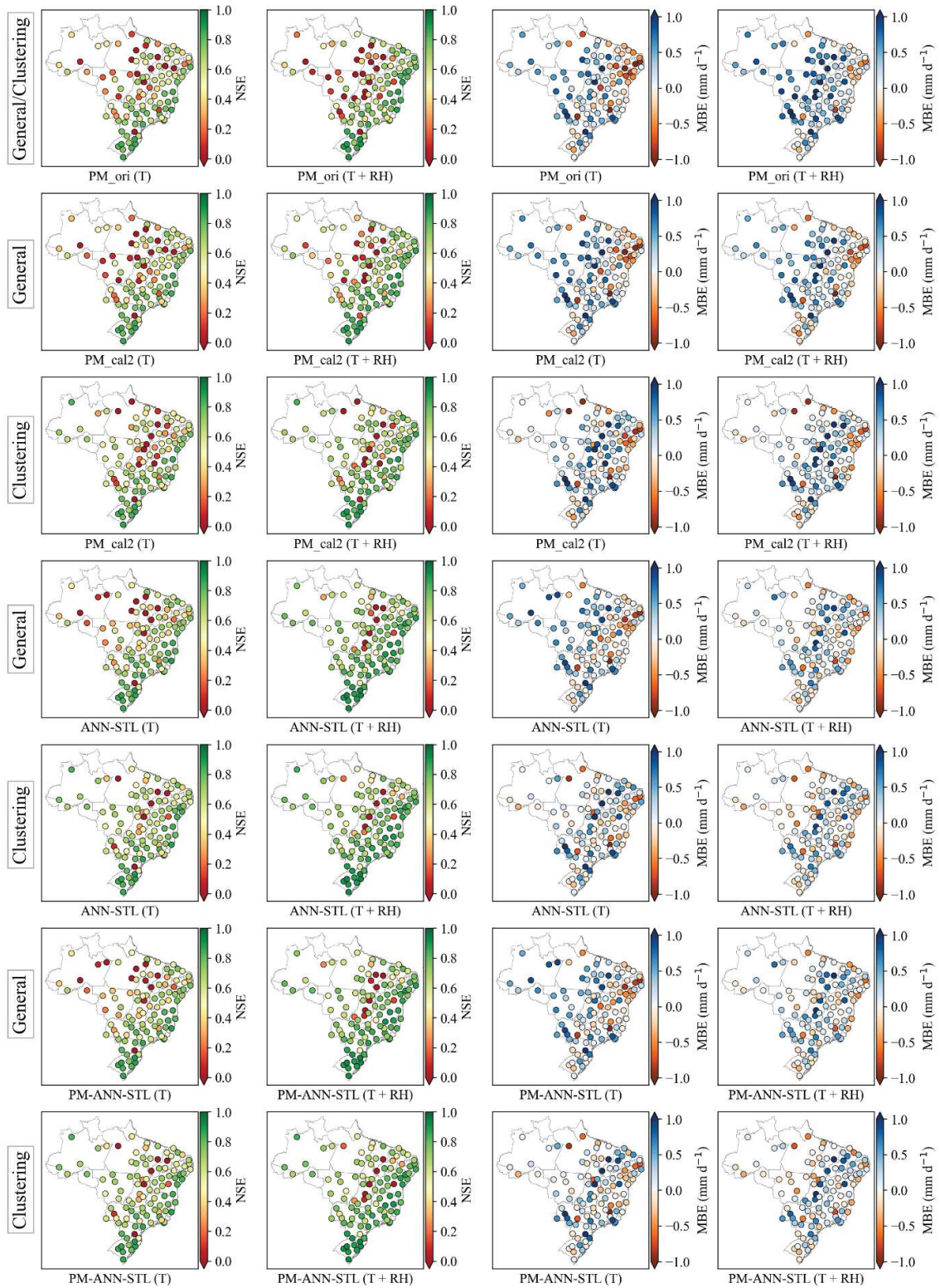
Although some studies have demonstrated performance gains by using MTL (Ng et al., 2019; Padarian et al., 2019), it does not promote gains in all cases, as found in the present study and also for Nunes et al. (2019). Thus, MTL should be considered as a tool with potential benefits for some tasks, but it needs to be evaluated for each particular case. Another way to explore possible advantages of MTL in the estimation of ETo, which can be assessed in future studies, is to add past and/or future ETo values (e.g., 1-day ahead ETo) as auxiliary target variables.

Overall, the combination between machine learning models and the FAO56-PM equation performed similarly to the machine learning models individually used in both general and clustering scenarios (Figs. 9, 10, 11 and 15). Using machine learning models to estimate missing data and the FAO56-PM equation to estimate ETo instead of using machine learning models to estimate ETo directly can make the ETo estimation process a little clearer. However, it requires the development and use of more models.

Although the MARS models used in the general scenario to estimate ETo and missing data were not so good as the ANN and RF models (Figs. 6, 7, 8, 9, 10), in contrast to most machine learning models, they can be used in the form of an algebraic mathematical equation, which can simplify its usage for an end user. In addition, the MARS models used individually or in combination with the FAO56-PM equation presented better performance to estimate ETo than the FAO56-PM equation with the same input data (Figs. 9, 10 and 11). In the temperature- and relative humidity-based setting, for the MARS model used individually, in relation to PM\_cal2, NSE increased from 0.56 to 0.65 (16.1%) and RMSE decreased from 0.74 to 0.67 (9.5%). For the temperature-based models, in relation to the PM\_ori model, the use of MARS provided only slight performance gains. To estimate solar radiation, actual vapor pressure/relative humidity and wind speed, the MARS models performed better than the FAO methodologies, mainly when using temperature and relative humidity as input (Figs. 6, 7 and 8). To make the MARS models developed in the present study available to the reader, they are provided (equations and an example spreadsheet) as supplementary data (Appendix A).

The distribution of NSE and MBE values over the Brazilian territory, for ETo estimated in the general and clustering scenarios, is presented in Fig. 16 for the FAO56-PM equation with

missing data estimated using traditional methodologies (original and the best calibrated version) and the ANN models developed using STL and their combinations with the FAO56-PM equation. Similarly, mean NSE values for each cluster defined using K-means are presented in Table 3 for the mentioned models and PM\_cal1. As previously observed, using relative humidity data in addition to temperature in the non-calibrated FAO56-PM equation (PM\_ori) resulted in worse ETo estimation (higher MBE values and lower NSE values) at some weather. However, it promoted better results (higher NSE values), in both general and clustering scenarios, when using the FAO56-PM equation calibrated and the machine learning models used individually and in combination with the FAO56-PM equation (Fig. 16 and Table 3).



**Fig. 16.** NSE and MBE values (for the test dataset) distribution over the Brazilian territory for different models used to estimate ETo in the general and clustering scenarios. T - temperature; RH - relative humidity; PM - FAO56-PM equation; PM\_ori - PM with missing data estimated using original FAO methodologies; PM\_cal2 - PM with missing data estimated using original

FAO methodologies and calibrated using linear regression; ANN - Artificial neural network; STL - single-task learning; MTL - multi-task learning.

**Table 3.** NSE values, for each cluster, referring to ETo estimated using the FAO56-PM equation with missing data and machine learning models developed in the general and clustering scenarios. Highlighted values indicate the best performing model for each cluster.

Scenario	Input type	Model	C1	C2	C3	C4	C5
General	T	PM_ori	0.37	0.57	0.70	0.30	<b>0.38</b>
		PM_cal1	0.24	0.56	0.65	0.31	0.18
		PM_cal2	0.31	0.57	0.69	0.28	0.33
		ANN-STL	0.36	0.61	<b>0.73</b>	<b>0.46</b>	0.25
		PM-ANN-STL	<b>0.38</b>	<b>0.62</b>	<b>0.73</b>	0.45	0.27
	T+RH	PM_ori	0.23	0.57	0.75	0.26	0.32
		PM_cal1	0.40	0.62	0.76	0.40	0.41
		PM_cal2	0.43	0.63	0.79	0.38	0.50
		ANN-STL	<b>0.59</b>	<b>0.67</b>	<b>0.85</b>	<b>0.57</b>	<b>0.66</b>
		PM-ANN-STL	0.58	<b>0.67</b>	<b>0.85</b>	0.55	<b>0.66</b>
Clustering	T	PM_ori	0.37	0.57	0.70	0.30	0.38
		PM_cal1	0.52	0.55	0.66	0.02	0.46
		PM_cal2	0.51	0.56	0.70	-0.01	0.45
		ANN-STL	0.60	<b>0.62</b>	<b>0.73</b>	0.31	0.60
		PM-ANN-STL	<b>0.61</b>	<b>0.62</b>	0.72	<b>0.32</b>	<b>0.61</b>
	T+RH	PM_ori	0.23	0.57	0.75	0.26	0.32
		PM_cal1	0.59	0.61	0.75	0.19	0.50
		PM_cal2	0.60	0.62	0.78	0.16	0.57
		ANN-STL	<b>0.65</b>	<b>0.69</b>	<b>0.85</b>	<b>0.51</b>	<b>0.68</b>
		PM-ANN-STL	0.64	0.67	<b>0.85</b>	0.49	<b>0.68</b>

C1, C2, C3, C4 and C5 stand for cluster 1 to 5, respectively; PM - FAO56-PM equation; PM\_ori - PM with missing data estimated using original FAO methodologies; PM\_cal1 - PM with missing data estimated using calibrated FAO methodologies; PM\_cal2 - PM with missing data estimated using original FAO methodologies and calibrated using linear regression; ANN - Artificial neural network; STL - single-task learning; NSE - Nash-Sutcliffe efficiency coefficient.

Overall, the temperature-based models from the general scenario performed poorly in northern Brazil (Fig. 16), which is associated with clusters 1 and 5 (Table 3). It is probably associated with the lower weather stations density in this region, which makes the models developed less adapted to this region. However, when considering the clustering approach, better performances were obtained. In this case, the models were able to better capture the relations between the input variables and ETo, since they were developed for a more homogeneous region. Adding relative humidity data also improved the results found in northern Brazil. In the general scenario, the use of relative humidity was an important factor in reducing ETo overestimation (high positive MBE values) in northern Brazil and reducing ETo underestimation (high negative MBE values) in northeastern Brazil, keeping MBE values closer to zero. Overall, the temperature- and relative humidity-based machine learning models and their combinations with the FAO56-PM equation provided the best performances, with higher NSE values and lower MBE absolute values, in both general and clustering scenarios. The best performances were found for clusters 2 and 3 in both general and clustering approaches. After the clustering, cluster 4 had its performance reduced in relation to the general scenario, which indicates that this cluster was not well defined for ETo estimation purposes.

Given the high climatic variability in the Brazilian territory, it is hard to develop general models to estimate ETo with good accuracy in any part of the country, especially when using temperature-based models. To achieve higher performances, future studies could investigate the use of extra information as input for the models, such as geographical coordinates, elevation, long-term gridded data obtained from ground data interpolation or satellite data, and other variables that can help the algorithms to achieve better performances. Defining regions with similar climatic conditions can also improve the performance of the models, as found in the present study. Future studies can better explore different approaches to apply clustering, such as choosing the best input variables and testing different clustering algorithms.

#### 4 Conclusions

In both general and clustering scenarios, using simple linear regression to calibrate the FAO56-PM equation provided slightly better results than calibrating the methodologies used to estimate missing data. Unexpectedly, the temperature-based FAO56-PM equation provided better ETo estimates before both calibration types assessed, which may be due to some random difference between data used to calibrate and test the equation. When using temperature and relative humidity data, better results were observed after both calibration types, reducing mean RMSE from 0.80 up to 0.74 mm d<sup>-1</sup> in the general and clustering scenarios.



Overall, the machine learning models performed better than the FAO56-PM equation to estimate ETo and better than the original and calibrated FAO methodologies to estimate missing data in both general and clustering scenarios. When combining machine learning models with the FAO56-PM equation, performance improvements were not observed, obtaining results similar to those found using machine learning models individually. Using multi-task learning (MTL) to develop ANN and RF models did not provide strong performance changes. Similar results were obtained in relation to using single-task learning (STL).

In the general scenario, for the temperature-based models, PM-ANN-STL performed the best to estimate ETo, increasing mean NSE from 0.49 to 0.53 (8.2%) and decreasing mean RMSE from 0.83 to 0.78 (6.0%), in relation to the non-calibrated FAO56-PM equation, which performed better than its calibrated versions. For the temperature- and relative humidity-based models, the ANN and RF models developed using STL and MTL performed the best, increasing NSE from 0.56 to 0.67 (19.6%) and decreasing RMSE from 0.74 to 0.65 (12.2%), in relation to the FAO56-PM equation calibrated using linear regression, which performed better than its other versions.

When dividing Brazil into five climatically homogeneous groups (clustering), performance gains were obtained in estimating ETo for the temperature-based machine learning models and the calibrated FAO56-PM equation. The best model (PM-ANN-STL) showed mean NSE and RMSE equal to 0.58 and 0.75 mm d<sup>-1</sup>, respectively. The temperature- and relative humidity-based models were little affected by the clustering. Similarly, better estimates of actual vapor pressure and wind speed were obtained after the clustering.

Among the machine learning models assessed, overall, ANN and RF performed the best, with a slight advantage for ANN. Although the MARS models showed inferior performance, they generally performed better than the original and calibrated FAO approaches to estimate ETo and missing data, especially when using temperature and relative humidity data. Thus, the MARS models developed are made available as supplementary data (Appendix A) (equations and an example spreadsheet).

## Acknowledgments

The present study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and CNPq, National Council for Scientific and Technological Development - Brazil. The authors wish to thank the Brazilian National Institute of Meteorology (INMET) for the meteorological data used.

## References

- Ahmadi, S.H., Javanbakht, Z., 2020. Assessing the physical and empirical reference evapotranspiration (ET<sub>o</sub>) models and time series analyses of the influencing weather variables on ET<sub>o</sub> in a semi-arid area. *J. Environ. Manage.* 276, 111278. doi:10.1016/j.jenvman.2020.111278
- Alencar, L.P. de, Sedyama, G.C., Mantovani, E.C., 2015. Estimativa da evapotranspiração de referência (ET<sub>o</sub> padrão FAO), para Minas Gerais, na ausência de alguns dados climáticos. *Eng. Agric.* 35, 39–50. (In Portuguese) doi:10.1590/1809-4430-Eng.Agric.v35n1p39-50/2015
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and Drainage, Paper No. 56, Food and Agriculture Organization of the United Nations, Rome.
- Almorox, J., Senatore, A., Quej, V.H., Mendicino, G., 2018. Worldwide assessment of the Penman–Monteith temperature approach for the estimation of monthly reference evapotranspiration. *Theor. Appl. Climatol.* 131, 693–703. doi:10.1007/s00704-016-1996-2
- Althoff, D., Bazame, H.C., Filgueiras, R., Dias, S.H.B., 2018. Heuristic methods applied in reference evapotranspiration modeling. *Cienc. e Agrotecnologia* 42, 314–324. doi:10.1590/1413-70542018423006818
- Althoff, D., Santos, R.A.D., Bazame, H.C., Cunha, F.F., Filgueiras, R., 2019. Improvement of Hargreaves–Samani reference evapotranspiration estimates with local calibration. *Water*, 11, 2272. doi:10.3390/w11112272
- Alvares, C.A., Stape, J.L., Sentelhas, P.C., De Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen’s climate classification map for Brazil. *Meteorol. Zeitschrift* 22, 711–728. doi:10.1127/0941-2948/2013/0507
- Awal, R., Habibi, H., Fares, A., Deb, S., 2020. Estimating reference crop evapotranspiration under limited climate data in West Texas. *J. Hydrol. Reg. Stud.* 28, 100677. doi:10.1016/j.ejrh.2020.100677
- Caruana, R., 1997. Multitask Learning. *Mach. Learn.* 28, 41–75. doi:10.1023/A:1007379606734
- Chen, T., Guestrin, C., 2016. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13-17-Aug.



- Cheng, M.Y., Cao, M.T., 2014. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Appl. Soft Comput. J.* 22, 178–188. doi:10.1016/j.asoc.2014.05.015
- Cunha, F.F. da, Magalhães, F.F., de Castro, M.A., de Souza, E.J., 2017. Performance of estimative models for daily reference evapotranspiration in the city of Cassilândia, Brazil. *Eng. Agric.* 37, 173–184. doi:10.1590/1809-4430-Eng.Agric.v37n1p173-184/2017
- De Bruin, H.A.R., Trigo, I.F., Bosveld, F.C., Meirink, J. F., 2016. A thermodynamically based model for actual evapotranspiration of an extensive grass field close to FAO reference, suitable for remote sensing application. *J. Hydrometeorol.* 17, 1373–1382. doi:10.1175/JHM-D-15-0006.1
- Djaman, K., O'Neill, M., Diop, L., Bodian, A., Allen, S., Koudahe, K., Lombard, K., 2019. Evaluation of the Penman-Monteith and other 34 reference evapotranspiration equations under limited data in a semiarid dry climate. *Theor. Appl. Climatol.* 137, 729–743. doi:10.1007/s00704-018-2624-0
- Exner-Kittridge, M., 2012. Closure to “Case Study on the Accuracy and Cost/Effectiveness in Simulating Reference Evapotranspiration in West-Central Florida” by Michael Grant Exner-Kittridge and Mark Cable Rains. *J. Hydrol. Eng.* 17, 225–226. doi:10.1061/(asce)he.1943-5584.0000430
- 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. doi:10.1016/j.compag.2017.01.027
- Ferreira, L.B., da Cunha, F.F., da Silva, G.H., Campos, F.B., Dias, S.H.B., Santos, J.E.O., 2021. Generalizability of machine learning models and empirical equations for the estimation of reference evapotranspiration from temperature in a semiarid region. *An. Acad. Bras. Cienc.* 93, e20200304. doi:10.1590/0001-3765202120200304
- Ferreira, L.B., da Cunha, F.F., de Oliveira, R.A., Fernandes Filho, E.I., 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. *J. Hydrol.* 572, 556–570. doi:10.1016/j.jhydrol.2019.03.028
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi:10.1002/joc.5086
- Hadria, R., Benabdelouhab, T., Lionboui, H., Salhi, A., 2021. Comparative assessment of different reference evapotranspiration models towards a fit calibration for arid and semi-arid areas. *J. Arid Environ.* 184, 104318. doi:10.1016/j.jaridenv.2020.104318

- He, C., Liu, J., Xu, F., Zhang, T., Chen, S., Sun, Z., Zheng, W., Wang, R., He, L., Feng, H., Yu, Q., He, J., 2020. Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods. *Energy Convers. Manag.* 220, 113111. doi:10.1016/j.enconman.2020.113111
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* 574, 1029–1041. doi:10.1016/j.jhydrol.2019.04.085
- Kiafar, H., Babazadeh, H., Marti, P., Kisi, O., Landaras, G., Karimi, S., Shiri, J., 2017. Evaluating the generalizability of GEP models for estimating reference evapotranspiration in distant humid and arid locations. *Theor. Appl. Climatol.* 130, 377–389. doi:10.1007/s00704-016-1888-5
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- 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. doi:10.1016/j.agrformet.2018.08.007
- Kodinariya, T.M., Makwana, P.R., 2013. Review on determining number of Cluster in K-Means Clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* 1, 2321–7782.
- Lindauer, M., Schmid, H.P., Grote, R., Steinbrecher, R., Mauder, M., Wolpert, B., 2017. A simple new model for incoming solar radiation dependent only on screen-level relative humidity. *J. Appl. Meteorol. Climatol.* doi:10.1175/JAMC-D-16-0085.1
- Mattar, M.A., Alazba, A.A., Alblewi, B., Gharabaghi, B., Yassin, M.A., 2016. Evaluating and Calibrating Reference Evapotranspiration Models Using Water Balance under Hyper-Arid Environment. *Water Resour. Manag.* 30, 3745–3767. doi:10.1007/s11269-016-1382-y
- Mehdizadeh, S., Behmanesh, J., Khalili, K., 2017. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput. Electron. Agric.* 139, 103–114. doi:10.1016/j.compag.2017.05.002
- National Water and Sanitation Agency (ANA), 2021. Atlas irrigação: uso da água na agricultura irrigada. 2ª edição. ANA, Brasília. (In Portuguese)
- Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma* 352, 251–267. doi:10.1016/j.geoderma.2019.06.016

- Nunes, M., Gerding, E., McGroarty, F., Niranjan, M., 2019. A comparison of multitask and single task learning with artificial neural networks for yield curve forecasting. *Expert Syst. Appl.* 119, 362–375. doi:10.1016/j.eswa.2018.11.012
- Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. *Geoderma Reg.* 16. doi:10.1016/j.geodrs.2018.e00198
- Paredes, P., Fontes, J.C., Azevedo, E.B., Pereira, L.S., 2018a. Daily reference crop evapotranspiration with reduced data sets in the humid environments of Azores islands using estimates of actual vapor pressure, solar radiation, and wind speed. *Theor. Appl. Climatol.* 134, 1115–1133. doi:10.1007/s00704-017-2329-9
- Paredes, P., Martins, D.S., Pereira, L.S., Cadima, J., Pires, C., 2018b. Accuracy of daily estimation of grass reference evapotranspiration using ERA-Interim reanalysis products with assessment of alternative bias correction schemes. *Agric. Water Manag.* 210, 340–353. doi:10.1016/j.agwat.2018.08.003
- Paredes, P., Pereira, L.S., 2019. Computing FAO56 reference grass evapotranspiration PM-ET<sub>o</sub> from temperature with focus on solar radiation. *Agric. Water Manag.* 215, 86–102. doi:10.1016/j.agwat.2018.12.014
- Pereira, L.S., Allen, R.G., Smith, M., Raes, D., 2015. Crop evapotranspiration estimation with FAO56: Past and future. *Agric. Water Manag.* doi:10.1016/j.agwat.2014.07.031
- Popova, Z., Kercheva, M., Pereira, L.S., 2006. Validation of the FAO methodology for computing ET<sub>o</sub> with limited data. Application to south Bulgaria. *Irrig. Drain.* 55, 201–215. doi:10.1002/ird.228
- Quej, V.H., Almorox, J., Ibrakhimov, M., Saito, L., 2016. Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico. *Energy Convers. Manag.* 110, 448–456. doi:10.1016/j.enconman.2015.12.050
- Reis, M.M., da Silva, A.J., Zullo Junior, J., Tuffi Santos, L.D., Azevedo, A.M., Lopes, É.M.G., 2019. Empirical and learning machine approaches to estimating reference evapotranspiration based on temperature data. *Comput. Electron. Agric.* 165. doi:10.1016/j.compag.2019.104937
- Ren, X., Qu, Z., Martins, D.S., Paredes, P., Pereira, L.S., 2016. Daily Reference Evapotranspiration for Hyper-Arid to Moist Sub-Humid Climates in Inner Mongolia, China: I. Assessing Temperature Methods and Spatial Variability. *Water Resour. Manag.* 30, 3769–3791. doi:10.1007/s11269-016-1384-9

- Tangune, B.F., Escobedo, J.F., 2018. Reference evapotranspiration in São Paulo State: Empirical methods and machine learning techniques. *Int. J. Water Resour. Environ. Eng.* 10, 33–44. doi:10.5897/IJWREE2018.0772
- Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* (Switzerland). doi:10.3390/w11050910
- Valiantzas, J.D., 2018a. Temperature-and humidity-based simplified Penman's ET<sub>0</sub> formulae. Comparisons with temperature-based Hargreaves-Samani and other methodologies. *Agric. Water Manag.* 208, 326–334. doi:10.1016/j.agwat.2018.06.028
- Valiantzas, J.D., 2018b. Modification of the Hargreaves–Samani Model for Estimating Solar Radiation from Temperature and Humidity Data. *J. Irrig. Drain. Eng.* doi:10.1061/(asce)ir.1943-4774.0001275
- Valiantzas, J.D., 2013. Simplified forms for the standardized FAO-56 Penman-Monteith reference evapotranspiration using limited weather data. *J. Hydrol.* 505, 13–23. doi:10.1016/j.jhydrol.2013.09.005
- Valiantzas, J.D., 2012. Discussion of “Case Study on the Accuracy and Cost/Effectiveness in Simulating Reference Evapotranspiration in West-Central Florida” by Michael Grant Exner-Kittridge and Mark Cable Rains. *J. Hydrol. Eng.* 17, 224–225. doi:10.1061/(asce)he.1943-5584.0000394
- 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. doi:10.1016/j.compag.2020.105653
- Zanetti, S.S., Dohler, R.E., Cecilio, R.A., Pezzopane, J.E.M., Xavier, A.C., 2019. Proposal for the use of daily thermal amplitude for the calibration of the Hargreaves-Samani equation. *J. Hydrol.* 571, 193–201. doi:10.1016/j.jhydrol.2019.01.049

## Appendix A. Supplementary data

### Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across Brazil

The multivariate adaptive regression splines (MARS) models developed to estimate reference evapotranspiration (ET<sub>o</sub>), solar radiation, maximum and minimum relative humidity and wind speed at 2 m height are provided in this appendix. As the data used to develop the models were normalized, before using the models, input variables should be normalized. Similarly, output variables predicted by the MARS models should be transformed to their normal scale. To normalize a particular input variable, the following equation should be used:

$$x_{ni} = \frac{x_i - \mu}{\sigma} \quad (1)$$

where  $x_{ni}$  is the normalized value,  $x_i$  is the original value,  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

To transform a particular output variable to its original scale, the following equation should be used:

$$y_i = (\sigma y_{ni}) + \mu \quad (1)$$

where  $y_i$  is the output value in original scale,  $y_{ni}$  is the normalized output value (output from MARS models),  $\mu$  is the mean, and  $\sigma$  is the standard deviation.

The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) referring to each possible input and output variables are presented in Table A1.

**Table A1.** Mean ( $\mu$ ) and standard deviation ( $\sigma$ ) values referring to each possible input and output variables used in the MARS models provided.

Input variables	$\mu$	$\sigma$
Maximum temperature (°C)(Tx)	29.0368	5.0586
Minimum temperature (°C)(Tn)	18.2566	4.6467
Maximum relative humidity (%) (RHx)	90.3330	8.9929
Minimum relative humidity (%) (RHn)	47.2278	16.8098
Extraterrestrial solar radiation (MJ m <sup>-2</sup> d <sup>-1</sup> ) (Ra)	34.4325	6.2931
Output variables	$\mu$	$\sigma$
Reference evapotranspiration (mm d <sup>-1</sup> ) (ETo)	4.1124	1.5575
Solar radiation (MJ m <sup>-2</sup> d <sup>-1</sup> ) (Rs)	18.2045	6.2286
Wind speed (m s <sup>-1</sup> ) (Ws)	1.6090	0.9842
Maximum relative humidity (%) (RHx)	90.3330	8.9929
Minimum relative humidity (%) (RHn)	47.2278	16.8098

## MARS MODELS

In the MARS models, a function  $h(x)$  is used, which is defined below:

$$h(x) = \text{Maximum}(0, x)$$

### Temperature-based models

Reference evapotranspiration (ETo)

$$\begin{aligned} ETo = & -0.0338 \cdot Tx^2 \cdot h(4.77 - Tn) - 0.25 \cdot Tx \cdot h(1.57 - Ra) - 0.022 \cdot Tx \cdot h(4.77 - Tn) + \\ & 84.9 \cdot Tx \cdot h(Tn - 4.77) + 1.38 \cdot Tx - 0.0595 \cdot Tn^2 \cdot h(4.77 - Tn) + 0.0171 \cdot Tn \cdot Ra^2 - 0.227 \cdot Tn \cdot Ra \\ & - 0.0605 \cdot Tn \cdot h(3.00 - Tx) \cdot h(4.77 - Tn) - 4.88 \cdot Tn \cdot h(4.77 - Tn) \cdot h(Tx - 3.00) - 0.121 \cdot Tn \cdot h(4.77 \\ & - Tn) - 50.4 \cdot Tn \cdot h(Tn - 4.77) + 0.96 \cdot Tn - 0.396 \cdot h(1.57 - Ra) + 1370.0 \cdot h(Ra - 1.57) + 0.609 \end{aligned}$$

Solar radiation (Rs)

$$\begin{aligned} Rs = & -0.326 \cdot Tx^2 + 0.0658 \cdot Tx \cdot Tn \cdot h(2.68 - Tn) + 0.173 \cdot Tx \cdot Tn - 0.251 \cdot Tx \cdot h(1.57 - Ra) + \\ & 1.3 \cdot Tx - 0.0768 \cdot Tn^2 \cdot h(2.68 - Tn) + 0.307 \cdot Tn \cdot h(1.57 - Ra) - 0.288 \cdot Tn \cdot h(2.68 - Tn) - \\ & 0.0401 \cdot Ra \cdot h(1.57 - Ra) - 0.504 \cdot h(1.57 - Ra) + 0.444 \cdot h(2.68 - Tn) - 2.23 \cdot h(Tn - 2.68) + \\ & 1800.0 \cdot h(Ra - 1.57) - 0.36 \end{aligned}$$

Maximum relative humidity (RHx)

$$RHx = 0.0379 \cdot Tx^2 \cdot h(2.98 - Tx) + 0.323 \cdot Tx \cdot h(2.98 - Tx) + 0.0385 \cdot Ra^2 - 0.066 \cdot Ra \cdot h(2.98 - Tx) + 0.385 \cdot Ra + 1.36 \cdot h(2.98 - Tx) + 35.0 \cdot h(Tx - 2.98) - 3.93$$

Minimum relative humidity (RHn)

$$RHn = -0.0112 \cdot Tx \cdot Ra \cdot h(2.68 - Tn) + 0.297 \cdot Tx \cdot h(2.68 - Tn) - 0.117 \cdot Tx \cdot h(2.98 - Tx) + 3.59 \cdot Tx \cdot h(Tx - 2.98) - 0.152 \cdot Tx \cdot h(Tn - 2.68) - 1.69 \cdot Tx + 0.108 \cdot Tn^2 - 0.0408 \cdot Tn \cdot Ra^2 + 0.0551 \cdot Tn \cdot Ra + 0.935 \cdot Tn + 0.0663 \cdot Ra - 0.0511$$

Wind speed at 2 m (Ws)

$$Ws = 0.0223 \cdot Tx^2 \cdot Ra + 0.0307 \cdot Tx^2 \cdot h(3.22 - Tn) - 0.0569 \cdot Tx^2 - 0.0567 \cdot Tx \cdot Tn \cdot h(3.22 - Tn) - 0.357 \cdot Tx \cdot h(3.22 - Tn) + 0.93 \cdot Tx + 0.144 \cdot Tn \cdot h(1.57 - Ra) - 0.0717 \cdot Ra \cdot h(1.57 - Ra) - 0.206 \cdot h(1.57 - Ra) + 0.201 \cdot h(3.22 - Tn) - 1.51 \cdot h(Tn - 3.22) + 2690.0 \cdot h(Ra - 1.57) - 0.449$$

### Temperature- and relative humidity-based models

Reference evapotranspiration (ETo)

$$ETo = 0.0171 \cdot Tx^2 \cdot RHn - 0.0068 \cdot Tx \cdot RHx \cdot RHn + 0.211 \cdot Tx \cdot RHn + 0.237 \cdot Tx - 0.0588 \cdot Tn \cdot RHx - 0.149 \cdot Tn \cdot RHn - 0.0902 \cdot RHx \cdot RHn - 0.263 \cdot RHx + 0.0734 \cdot RHn^2 + 0.168 \cdot RHn \cdot h(1.57 - Ra) - 0.7 \cdot RHn - 0.457 \cdot h(1.57 - Ra) + 724.0 \cdot h(Ra - 1.57) + 0.805$$

Solar radiation (Rs)

$$Rs = 0.0365 \cdot Tx \cdot RHn \cdot h(1.57 - Ra) - 0.0701 \cdot Tx \cdot RHn \cdot h(2.33 - Tn) - 0.173 \cdot Tx \cdot RHn \cdot h(Tn - 2.33) + 0.329 \cdot Tx \cdot RHn + 0.0196 \cdot RHn^3 - 0.0164 \cdot RHn^2 + 0.211 \cdot RHn \cdot h(1.57 - Ra) - 1.01 \cdot RHn - 0.454 \cdot h(1.57 - Ra) + 1230.0 \cdot h(Ra - 1.57) + 0.844$$

Wind speed at 2 m (Ws)

$$Ws = -0.723 \cdot Tx + 0.0831 \cdot Tn \cdot h(1.57 - Ra) + 0.143 \cdot Tn - 0.0375 \cdot RHx^2 - 0.137 \cdot RHx \cdot RHn + 0.067 \cdot RHx \cdot h(2.33 - Tn) - 0.598 \cdot RHx + 0.0409 \cdot RHn^2 + 0.132 \cdot RHn \cdot h(2.33 - Tn) + 0.405 \cdot RHn \cdot h(Tn - 2.33) - 0.59 \cdot RHn - 0.0404 \cdot Ra \cdot h(1.57 - Ra) - 0.238 \cdot h(1.57 - Ra) + 1660.0 \cdot h(Ra - 1.57) + 0.445$$



## Article 4: Selecting models for the estimation of reference evapotranspiration for irrigation scheduling purposes

**Abstract:** Alternative models for the estimation of reference evapotranspiration (ET<sub>o</sub>) are typically assessed using traditional error metrics, such as root mean square error (RMSE), which may not be sufficient to select the best model for irrigation scheduling purposes. Thus, this study analyzes the performance of the original and calibrated Hargreaves-Samani (HS), Romanenko (ROM) and Jensen-Haise (JH) equations, initially assessed using traditional error metrics, for use in irrigation scheduling, considering the simulation of different irrigation intervals/time scales. Irrigation scheduling was simulated using meteorological data collected in Viçosa-MG and Mocambinho-MG, Brazil. The Penman-Monteith FAO-56 equation was used as benchmark. In general, the original equations did not perform well to estimate ET<sub>o</sub>, except the ROM and HS equations used at Viçosa and Mocambinho, respectively. Calibration and the increase in the time scale provided performance gains. When applied in irrigation scheduling, the calibrated HS and JH equations showed the best performances. Even with greater errors in estimating ET<sub>o</sub>, the calibrated HS equation performed similarly or better than the calibrated JH equation, as it had errors with greater potential to be canceled during the soil water balance. Finally, in addition to using error metrics, the performance of the models throughout the year should be considered in their assessment. Furthermore, simulating the application of ET<sub>o</sub> models in irrigation scheduling can provide valuable information for choosing the most suitable model.

**Keywords:** calibration, empirical equations, error metrics, model assessment

### 1 Introduction

Irrigation is a very important practice to ensure good agricultural productions in arid and semiarid areas. In addition, it can contribute to reduce production risks, even in areas with reasonable rainfall levels, and can be used in greenhouse production. However, despite its benefits, irrigation should be used properly to avoid excessive or insufficient water application. In this sense, irrigation scheduling plays a key role, allowing one to provide water to different crops according to their requirements [1].

Irrigation scheduling can be performed using different approaches, but it is commonly based on reference evapotranspiration (ET<sub>o</sub>), which is typically computed using meteorological data [2–7]. ET<sub>o</sub> can be used as basis to compute the evapotranspiration of different crops. To accomplish this, a crop coefficient (K<sub>c</sub>) and a water stress coefficient (K<sub>s</sub>) are used to convert

ETo to the evapotranspiration of a particular crop, considering its development phase and the soil water availability [6,8].

ETo can be estimated using the Penman-Monteith FAO-56 (PM) equation, recommended by the Food and Agriculture Organization (FAO) [2,8]. This equation performs well in different regions of the world. However, in places with low meteorological data availability, its application becomes limited, since it requires air temperature, relative humidity, solar radiation and wind speed data [9,10].

To make it possible to estimate ETo using fewer meteorological data, several studies have evaluated the potential of empirical equations and machine learning models to estimate ETo under different meteorological data availability scenarios [10–15]. These alternative models can be important options for the estimation of ETo, however they typically have a limited performance. According to the performance of a particular model, it can be considered suitable or not for irrigation scheduling purposes.

To assess the performance of models for the estimation of ETo, traditional error metrics, such as root mean square error (RMSE), mean absolute error (MAE), mean bias error (MBE) and coefficient of determination ( $R^2$ ), are typically used [11–14,16]. Overall, these metrics compute the dissimilarity (error) or similarity between the estimates provided by a reference model, which is commonly represented by the PM equation, and a model under evaluation. Based on a single error metric or on a set of error metrics, it is possible to define the most efficient model to estimate ETo as the one with lower errors in relation to the reference model. However, when selecting models for irrigation scheduling, the use of the strategy mentioned above do not provide a direct assessment of the performance of the models for this specific purpose.

In irrigation scheduling, irrigation frequency can have a significant influence on the performance of the models since when grouping daily ETo values in longer periods, the prediction errors may decrease. In addition, when calculating crop evapotranspiration (ETc) using a water stress coefficient ( $K_s$ ), problems with ETo overestimation, which cause ETc overestimation, can be partially reduced during the soil water balance since the estimated soil water content will drop faster, promoting higher  $K_s$  reduction, which reduces the next ETc values calculated. Other important factor is the behavior of the ETo model over time. For instance, a model with random errors over time can has its errors partially canceled during the soil water balance. Finally, the rainfall distribution over the year can also impact the performance of irrigation scheduling performed with alternative ETo models. Given the

dynamics of irrigation scheduling, it is highlighted that the simple use of error metrics may not be sufficient to select the best ETo model for irrigation scheduling purposes.

Despite the importance of the development of methodologies for a better assessment of models for the estimation of ETo for irrigation scheduling purposes, according to our knowledge, so far, this type of study has not been found. Thus, the objective of this study was to analyze the performance of three original and calibrated empirical equations, initially evaluated using traditional error metrics, for irrigation scheduling, considering the simulation of different irrigation intervals.

## 2 Materials and methods

### Database

Hourly data from two automatic weather stations (2015-2017) of the Brazilian National Institute of Meteorology (INMET) located in the municipalities of Viçosa and Mocimbinho, which are located in the state of Minas Gerais, Brazil, were used. Maximum and minimum air temperature, mean relative humidity, solar radiation, wind speed (10 m) and rainfall data were used. Wind speed measured at 10 m height was converted to 2 m height, as suggested by Allen et al. [8]. The hourly data were converted to a daily timescale. Days with missing data were removed. The weather stations used in this study were selected because they represent relatively different climatic conditions. The mean values of the meteorological variables used, in the periods considered to calibrate the equations (2015-2016) and to assess their performances (2017), are presented in Table 1. The database is available in Supporting information or directly from INMET (<https://portal.inmet.gov.br/dadoshistoricos>).

**Table 1.** Mean values of the meteorological variables used in the study.

Viçosa (Latitude: -20.77°, longitude: -42.87° and altitude: 712 m)							
Period	T <sub>max</sub>	T <sub>min</sub>	RH	Ws	Rs	P	ETo
2015-2016	28.1	16.2	78.1	0.7	16.7	1194	3.3
2017	27.4	15.4	76.7	0.6	16.5	847	3.2
Mocimbinho (Latitude: -15.08°, longitude: -44.00° and altitude: 460 m)							
2015-2016	33.6	19.1	56.7	0.8	21.6	551	4.6
2017	32.9	18.3	57.1	0.9	21.9	573	4.6

T<sub>max</sub> - maximum air temperature (°C); T<sub>min</sub> - minimum air temperature (°C); RH - mean relative humidity (%); Ws - wind speed (2 m) (m s<sup>-1</sup>); Rs - solar radiation (MJ m<sup>-2</sup> d<sup>-1</sup>); P - annual rainfall (mm); ETo - reference evapotranspiration (mm d<sup>-1</sup>).

### Irrigation scheduling - Simulation configurations

To carry out irrigation scheduling, the soil water inputs (rainfall and irrigation) and output (evapotranspiration) were computed. Crop evapotranspiration (ET<sub>c</sub>) was calculated based on Equation 1, as recommended by Allen et al. [8] and Bernardo et al. [17]. K<sub>s</sub> coefficient is used to adjust ET<sub>c</sub> for water deficit conditions. When adjusted for water deficit conditions, as considered in the present study, it is common to refer to ET<sub>c</sub> as actual evapotranspiration (ET<sub>a</sub>) or adjusted ET<sub>c</sub>. In this study, the denotation ET<sub>c</sub> was maintained.

$$ET_c = ETo K_c K_s \quad (1)$$

where ET<sub>c</sub> - crop evapotranspiration, mm d<sup>-1</sup>; ETo - reference evapotranspiration, mm d<sup>-1</sup>; K<sub>c</sub> - crop coefficient; K<sub>s</sub> - water stress coefficient.

ETo was obtained using different equations, which are presented later. K<sub>s</sub> was calculated based on Equation 2 [17].

$$K_s = \frac{\ln(SWC+1)}{\ln(TAW+1)} \quad (2)$$

Where SWC - soil water content, mm; TAW - total available water, mm.

$$TAW = \frac{(FC-PWP)}{10} BD z \quad (3)$$

Where TAW - total available water, mm; FC - field capacity, % (water mass over dry soil mass); PWP - permanent wilting point, % (water mass over dry soil mass); BD - soil bulk density, g cm<sup>-3</sup>; z - effective rooting depth, cm.

Once ET<sub>c</sub> has been obtained, the soil water balance was computed based on Equation 4. The initial value of the soil water content (SWC) was equal to TAW. Effective rainfall (rainfall stored in the root zone) was considered equal to total rainfall, if total rainfall does not exceed the current soil water deficit (TAW – SWC), or equal to the current soil water deficit, otherwise.

$$SWC_i = SWC_{i-1} - ET_c + Pe + I \quad (4)$$

Where SWC<sub>i</sub> - soil water content on the current day, mm; SWC<sub>i-1</sub> - soil water content on the previous day, mm; ET<sub>c</sub> - crop evapotranspiration, mm; Pe - effective rainfall, mm; I - net irrigation depth, mm.

Knowing the current SWC, irrigation was computed in order to return SWC to field capacity. Thus, net irrigation depth was obtained by subtracting SWC from TAW (TAW - SWC). The parameters used for the simulations were as follows: field capacity (FC) = 30%, permanent wilting point (PWP) = 15%, soil bulk density (BD) = 1.1 g cm<sup>-3</sup>, effective rooting depth (z) = 20 cm, and crop coefficient (K<sub>c</sub>) = 1.1. Fixed irrigation intervals (1, 2, 4, 6 and 8

days) and variable irrigation intervals were considered. For variable irrigation intervals, the critical minimum soil water content was defined as 50% of TAW, which is considered by using a soil water depletion fraction for no stress ( $p$ ), also called soil water availability factor ( $f$ ), equal to 0.5. It is assumed that below this water content the crop begins to be affected by water deficit. To prevent the soil water content from exceeding the aforementioned critical minimum limit, irrigation was carried out when the soil water content was 40% below TAW. The simulations were performed using data from the year 2017, with data from 2015-2016 reserved to calibrate the empirical equations.

### Estimation of reference evapotranspiration

Daily ETo estimated using the PM equation (Equation 5) was employed as the standard method for calibration and evaluation of the empirical equations. All procedures necessary to calculate ETo were performed according to the recommendations of Allen et al. [8]. Although the PM equation is also subject to errors, it has good reliability and can be used as a standard for the development and calibration of other models [8,9].

$$ETo = \frac{0.408 \Delta (R_n - G) + \gamma \frac{900}{T_{mean} + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 u_2)} \quad (5)$$

where ETo - reference evapotranspiration, mm d<sup>-1</sup>;  $R_n$  - net solar radiation, MJ m<sup>-2</sup> day<sup>-1</sup>;  $G$  - soil heat flux, MJ m<sup>-2</sup> day<sup>-1</sup> (considered to be null for daily estimates);  $T_{mean}$  - daily mean air temperature, °C;  $u_2$  - wind speed at a 2 m height, m s<sup>-1</sup>;  $e_s$  - saturation vapor pressure, kPa;  $e_a$  - actual vapor pressure, kPa;  $\Delta$  - slope of the saturation vapor pressure function, kPa °C<sup>-1</sup>; and  $\gamma$  - psychrometric constant, kPa °C<sup>-1</sup>.

ETo was also estimated using the empirical equations shown in Table 2.

**Table 2.** Empirical equations used in the study.

Name / Inputs	Equation	Reference
Hargreaves-Samani (T)	$ETo = 0.0023 R_a (T_{mean} + 17.8) (T_{max} - T_{min})^{0.5}$	[18]
Romanenko (T, RH)	$ETo = 0.00006(25 + T_{mean})^2(100 - RH)$	[19]
Jensen-Haise (T, $R_s$ )	$ETo = 0.408 R_s (0.0252 T_{mean} + 0.078)$	[20]

T - air temperature, °C; RH – mean relative humidity, %;  $R_a$  - extraterrestrial radiation, mm d<sup>-1</sup>;  $T_{max}$  - maximum air temperature, °C;  $T_{min}$  - minimum air temperature, °C;  $T_{mean}$  - mean air temperature ( $[T_{max} + T_{min}]/2$ ), °C;  $R_s$  - solar radiation, MJ m<sup>-2</sup> d<sup>-1</sup>.

To adjust the empirical equations to the local climate conditions, they were calibrated based on simple linear regression, as recommended by Allen et al. [8], using data from 2015 to

2016. For this, daily ETo values estimated by the equation to be calibrated were used as the independent variable and ETo values estimated by the PM equation were used as the dependent variable. The intercept (a) and slope (b) values were used as calibration parameters, according to the following equation. The values obtained for the calibration parameters “a” and “b” are presented in Table 3.

$$ETo_{cal} = a + b(ETo) \quad (6)$$

where  $ETo_{cal}$  - calibrated reference evapotranspiration, mm d<sup>-1</sup>; “a” and “b” - calibration parameters; ETo - reference evapotranspiration estimated by the original (non-calibrated) empirical equation, mm d<sup>-1</sup>.

**Table 3.** Calibration parameters obtained for the different empirical equations evaluated at Viçosa and Mocambinho stations.

Equation	Viçosa		Mocambinho	
	a	b	a	b
Hargreaves-Samani	-0.76	0.90	-0.46	0.92
Romanenko	0.79	0.83	2.26	0.33
Jensen-Haise	0.35	0.66	0.51	0.61

### Performance comparison criteria

The performance of the empirical equations for the estimation of ETo was evaluated using data from the year 2017, the same period considered for irrigation scheduling. For that, ETo obtained with the PM equation was used as standard. The empirical equations were evaluated in different time scales (1, 2, 4, 6 and 8 days) by summing daily estimates. The error metrics listed below were used. Except for coefficient of determination ( $R^2$ ), normalized values of each error metric were calculated. For that, the error metrics values were divided by the mean of the analyzed variable (mean of the observed values). For time scales equal or greater than 2 days, the error metrics, except for  $R^2$ , were divided by the time scale in order to keep the unit mm d<sup>-1</sup>.

$$RMSE = \sqrt{\frac{1}{n} \sum (P_i - O_i)^2} \quad (7)$$

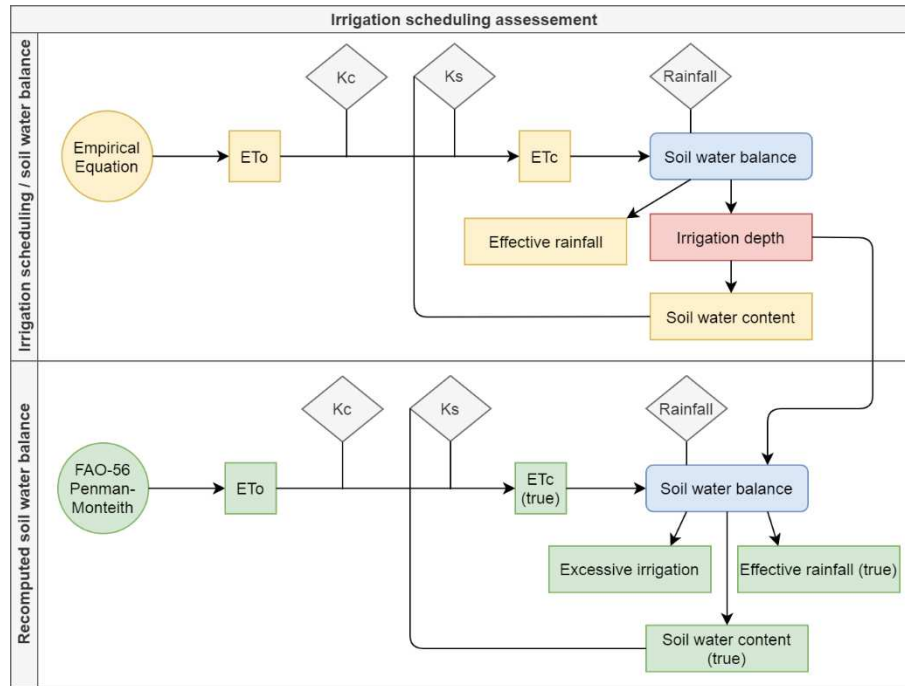
$$MAE = \frac{1}{n} \sum |P_i - O_i| \quad (8)$$

$$MBE = \frac{1}{n} \sum (P_i - O_i) \quad (9)$$

$$R^2 = \left[ \frac{\sum (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{(\sum (P_i - \bar{P})^2)(\sum (O_i - \bar{O})^2)}} \right]^2 \quad (10)$$

where RMSE - root mean square error, mm d<sup>-1</sup>; MAE - mean absolute error, mm d<sup>-1</sup>; MBE - mean bias error, mm d<sup>-1</sup>; R<sup>2</sup> - coefficient of determination; P<sub>i</sub> - predicted value, mm d<sup>-1</sup>; O<sub>i</sub> - observed value, mm d<sup>-1</sup>;  $\bar{P}$  - mean of the predicted values, mm d<sup>-1</sup>;  $\bar{O}$  - mean of the observed values, mm d<sup>-1</sup>; n - number of data pairs.

To assess the performance of the equations in the simulated irrigation scheduling, total ET<sub>c</sub>, total net irrigation depth and total effective rainfall estimated when using each equation were compared. In addition, after the end of the irrigation scheduling carried out with each empirical equation, the soil water balance was recomputed considering the irrigations recommended over the management period and ET<sub>c</sub> recalculated using ET<sub>o</sub> obtained with the PM equation (Fig 1). All the computations were performed in a daily basis. ET<sub>c</sub> and effective rainfall obtained in the recomputed soil water balance were denoted as ET<sub>c</sub> (true) and Pe (true), respectively. This procedure was performed to assess the real performance of the irrigation scheduling carried out with the different empirical equations. In this way, it is possible to analyze ET<sub>c</sub> that actually occurred during the management period and check the occurrence of irrigation excesses or deficits when using the different empirical equations to schedule irrigation.



**Fig 1.** Proposed methodology to assess the performance of different empirical equations for irrigation scheduling.



Based on the recomputed soil water balance (Fig 1), the irrigation excesses and deficits occurred during the management period were calculated. Irrigation excesses were computed as the sum of the net irrigation depths that resulted in soil water contents that exceeded field capacity. To compute irrigation deficits for the simulations with fixed irrigation intervals, deficit was considered as the reduction of ET<sub>c</sub> (true) observed for each empirical equation in relation to ET<sub>c</sub> observed when using the PM equation to schedule irrigation. It was done because water deficit promotes reductions in ET<sub>c</sub>, which is related to a worse crop development. For irrigation scheduling with variable irrigation intervals, the occurrence of deficits was computed as the sum of the soil water content deficits in relation to the critical minimum water content considered (50% of TAW,  $f=0.5$ ). Two classes of deficit were defined: (i) cases in which deficits were equivalent to  $0.5 < f \leq 0.6$  (weak deficit), and (ii)  $f > 0.6$  (moderate to strong deficit). These deficits were calculated according to Equations 11 and 12.

$$\text{Deficit}(0.5 < f \leq 0.6) = \sum (CL_1 - SWC_i), \text{ for } CL_2 \leq SWC_i < CL_1 \quad (11)$$

$$\text{Deficit}(f > 0.6) = \sum (CL_1 - SWC_i), \text{ for } SWC_i < CL_2 \quad (12)$$

$$CL_1 = 0.5TAW \quad (13)$$

$$CL_2 = 0.4TAW \quad (14)$$

where  $CL_1$  - critical limit soil water content referring to  $f=0.5$ , mm;  $CL_2$  - critical limit soil water content referring to  $f=0.6$ , mm;  $SWC_i$  - soil water content value, mm; TAW - total available water, mm.

### 3 Results and discussion

#### Estimation of ET<sub>o</sub>

Among the non-calibrated equations, the Romanenko equation (ROM) had the best performance for the estimation of ET<sub>o</sub> at Viçosa, with lower RMSE and MAE values in the various time scales considered (Table 4). This equation was followed by the Jensen-Haise (JH) equation and the Hargreaves-Samani (HS) equation, in that order. However, after calibration, the ROM equation exhibited the worst performance. The best performance was obtained by the JH equation, followed by the HS equation.

**Table 4.** Performance of the original and calibrated HS, ROM and JH equations for different time scales at Viçosa. Values in parentheses indicate the normalized error metrics.

Equation	Scale (d)	RMSE (mm d <sup>-1</sup> )	MAE (mm d <sup>-1</sup> )	MBE (mm d <sup>-1</sup> )	R <sup>2</sup>
HS	1	1.30 (41%)	1.21 (38%)	1.20 (38%)	0.84
	2	1.26 (40%)	1.20 (38%)	1.20 (38%)	0.89
	4	1.24 (39%)	1.20 (38%)	1.20 (38%)	0.92
	6	1.22 (39%)	1.19 (38%)	1.19 (38%)	0.94
	8	1.22 (39%)	1.20 (38%)	1.20 (38%)	0.96
HS_cal	1	0.47 (15%)	0.36 (11%)	0.01 (0%)	0.84
	2	0.37 (12%)	0.28 (9%)	0.01 (0%)	0.89
	4	0.29 (9%)	0.21 (7%)	0.00 (0%)	0.92
	6	0.25 (8%)	0.19 (6%)	0.00 (0%)	0.94
	8	0.18 (6%)	0.13 (4%)	0.00 (0%)	0.96
ROM	1	0.72 (23%)	0.56 (18%)	-0.10 (-3%)	0.68
	2	0.65 (20%)	0.52 (16%)	-0.09 (-3%)	0.70
	4	0.60 (19%)	0.50 (16%)	-0.09 (-3%)	0.69
	6	0.59 (19%)	0.48 (15%)	-0.09 (-3%)	0.67
	8	0.58 (18%)	0.49 (15%)	-0.09 (-3%)	0.66
ROM_cal	1	0.70 (22%)	0.59 (19%)	0.17 (6%)	0.68
	2	0.64 (20%)	0.55 (17%)	0.18 (6%)	0.70
	4	0.60 (19%)	0.53 (17%)	0.18 (6%)	0.69
	6	0.59 (19%)	0.53 (17%)	0.18 (6%)	0.67
	8	0.58 (18%)	0.52 (17%)	0.18 (6%)	0.66
JH	1	1.23 (39%)	1.09 (34%)	1.08 (34%)	0.97
	2	1.20 (38%)	1.08 (34%)	1.08 (34%)	0.98
	4	1.17 (37%)	1.08 (34%)	1.08 (34%)	0.98
	6	1.16 (37%)	1.08 (34%)	1.08 (34%)	0.98
	8	1.15 (36%)	1.08 (34%)	1.08 (34%)	0.98
JH_cal	1	0.20 (6%)	0.15 (5%)	-0.02 (-1%)	0.97
	2	0.18 (6%)	0.14 (5%)	-0.02 (-1%)	0.98
	4	0.17 (5%)	0.14 (4%)	-0.02 (-1%)	0.98
	6	0.17 (5%)	0.14 (4%)	-0.02 (-1%)	0.98
	8	0.16 (5%)	0.14 (4%)	-0.02 (-1%)	0.98

HS - Hargreaves-Samani; ROM - Romanenko; JH - Jensen-Haise. “\_cal” indicates the calibrated version of an equation.

At Mocambinho, the HS equation showed the best performance among the non-calibrated equations, followed by the JH and ROM equations, in that order (Table 5). After calibration, as for Viçosa, the JH equation showed the best performance, followed by the HS and ROM equations. Possibly the HS equation obtained the best performance among the non-calibrated equations because it was developed for a dry climate region (semiarid) [21], such as Mocambinho.

**Table 5.** Performance of the original and calibrated HS, ROM and JH equations for different time scales at Mocambinho. Values in parentheses indicate the normalized error metrics.

Equation	Scale (d)	RMSE (mm d <sup>-1</sup> )	MAE (mm d <sup>-1</sup> )	MBE (mm d <sup>-1</sup> )	R <sup>2</sup>
HS	1	1.00 (22%)	0.87 (19%)	0.82 (18%)	0.76
	2	0.95 (21%)	0.83 (18%)	0.82 (18%)	0.81
	4	0.90 (20%)	0.82 (18%)	0.82 (18%)	0.86
	6	0.89 (19%)	0.82 (18%)	0.82 (18%)	0.88
	8	0.88 (19%)	0.82 (18%)	0.82 (18%)	0.88
HS_cal	1	0.57 (13%)	0.43 (9%)	-0.07 (-2%)	0.76
	2	0.48 (11%)	0.37 (8%)	-0.07 (-2%)	0.81
	4	0.39 (9%)	0.32 (7%)	-0.07 (-2%)	0.86
	6	0.35 (8%)	0.28 (6%)	-0.07 (-2%)	0.88
	8	0.33 (7%)	0.27 (6%)	-0.07 (-2%)	0.88
ROM	1	2.55 (56%)	2.19 (48%)	2.04 (45%)	0.39
	2	2.52 (55%)	2.19 (48%)	2.05 (45%)	0.39
	4	2.49 (55%)	2.15 (47%)	2.05 (45%)	0.38
	6	2.45 (54%)	2.13 (47%)	2.05 (45%)	0.36
	8	2.43 (53%)	2.12 (46%)	2.05 (45%)	0.34
ROM_cal	1	0.90 (20%)	0.77 (17%)	-0.11 (-2%)	0.39
	2	0.85 (19%)	0.73 (16%)	-0.11 (-2%)	0.39
	4	0.80 (18%)	0.69 (15%)	-0.11 (-2%)	0.38
	6	0.79 (17%)	0.67 (15%)	-0.11 (-2%)	0.36
	8	0.78 (17%)	0.67 (15%)	-0.11 (-2%)	0.34

JH	1	2.10 (46%)	1.94 (43%)	1.94 (43%)	0.91
	2	2.07 (46%)	1.94 (43%)	1.94 (43%)	0.91
	4	2.05 (45%)	1.94 (43%)	1.94 (43%)	0.92
	6	2.04 (45%)	1.94 (43%)	1.94 (43%)	0.92
	8	2.03 (45%)	1.94 (43%)	1.94 (43%)	0.92
JH_cal	1	0.36 (8%)	0.26 (6%)	-0.07 (-2%)	0.91
	2	0.33 (7%)	0.25 (5%)	-0.07 (-2%)	0.91
	4	0.30 (7%)	0.23 (5%)	-0.07 (-2%)	0.92
	6	0.28 (6%)	0.21 (5%)	-0.07 (-2%)	0.92
	8	0.28 (6%)	0.20 (4%)	-0.07 (-2%)	0.92

HS - Hargreaves-Samani; ROM - Romanenko; JH - Jensen-Haise. “\_cal” indicates the calibrated version of an equation.

By increasing the time scale, there were performance gains for all the equations at both municipalities considered, with reductions in the error metrics (RMSE, MAE and MBE) and increase in  $R^2$ . This is because part of the errors in daily estimates can be canceled when considering longer time periods.

All the non-calibrated equations evaluated, with exception for the ROM equation used at Viçosa, showed relatively high MBE values at both studied locations, which indicates that there was a systemic overestimation of ETo. These equations obtained only small performance gains with the increase of the time scale. Furthermore, they did not reach RMSE and MAE values as low as those obtained by the calibrated equations, which showed a low general tendency to overestimate or underestimate ETo (low MBE absolute values).

Calibration promoted large reductions in RMSE and MAE values. After calibration, the equations with higher  $R^2$  values, with emphasis on the JH equation, even with high RMSE and MAE values before calibration, exhibited low errors. It should be noted that equations with good structure, which can adequately map the relationship between the input and output variables, reaching high  $R^2$  values, can be benefited by calibration [16].

Based on the metrics presented in Tables 4 and 5, one can easily rank the performance of the models, identifying those with the highest performances. However, it can still be difficult to infer whether a particular model is suitable or not for irrigation scheduling purposes.

### **Irrigation scheduling**

The results of the irrigation scheduling simulations with fixed irrigation intervals for Viçosa and Mocambinho are shown in Tables 6 and 7, respectively. The increase in irrigation

intervals promoted, in all cases, reductions in ETc values and in the total net irrigation depths applied. The decrease in ETc occurs due to the larger reductions in the soil moisture promoted by larger irrigation intervals, which reduces Ks values and, consequently, ETc. The reduction in the net irrigation depths occurs due to the reduction in ETc and due to the increase in effective rainfall, as seen in Tables 6 and 7. Longer irrigation intervals promote greater use of rainfall (i.e., more rainwater is stored in the root zone) because they increase the chance of soil having less moisture, in relation to shorter irrigation intervals, when rainfall reaches the soil.

**Table 6.** Information on the irrigation scheduling carried out at Viçosa with the PM equation and original and calibrated empirical equations considering different irrigation intervals (II). All the variables, except for II, are expressed in mm.

Equation	II (d)	ETc	ETc (true)	NID	Pe	Pe (true)	Deficit	Excess
PM	1	1204	1204	1014	189	189	-	-
	2	1186	1186	958	229	229	-	-
	4	1150	1150	863	286	286	-	-
	6	1114	1114	767	348	348	-	-
	8	1074	1074	697	377	377	-	-
HS	1	1663	1204	1402	262	189	0	387
	2	1631	1186	1318	314	229	0	360
	4	1561	1150	1187	374	286	0	324
	6	1485	1114	1059	426	348	0	292
	8	1393	1074	930	463	377	0	233
HS_cal	1	1206	1192	999	207	214	12	22
	2	1189	1175	941	248	253	11	19
	4	1154	1138	849	305	304	11	14
	6	1120	1105	759	361	358	9	12
	8	1082	1063	686	396	385	11	8
ROM	1	1166	1184	1005	160	256	20	77
	2	1149	1168	947	201	289	18	69
	4	1113	1132	853	260	340	17	60
	6	1077	1094	761	316	387	20	53
	8	1035	1054	692	343	407	20	45
ROM_cal	1	1270	1191	1088	182	232	13	129

	2	1251	1174	1027	224	270	12	122
	4	1211	1139	928	282	321	11	111
	6	1170	1101	828	342	373	13	100
	8	1122	1061	757	365	398	13	93
JH	1	1614	1204	1391	223	189	0	377
	2	1581	1186	1309	271	229	0	352
	4	1508	1150	1166	342	286	0	303
	6	1425	1114	1029	395	348	0	263
	8	1331	1074	899	431	377	0	202
JH_cal	1	1196	1197	1011	185	209	6	23
	2	1179	1180	955	224	247	6	22
	4	1143	1144	864	280	300	6	20
	6	1109	1109	769	340	359	5	19
	8	1070	1068	700	370	385	6	17

NID - total net irrigation depth; Pe - effective rainfall; Deficit - ETc deficit in relation to ETc obtained in the irrigation scheduling performed with the PM equation; Excess - excessive irrigation; ETc (true) and Pe (true) - ETc and Pe recalculated using ETo obtained with the PM equation. All the variables are expressed in mm. HS - Hargreaves-Samani; ROM - Romanenko; JH - Jensen-Haise. “\_cal” indicates the calibrated version of an equation.

**Table 7.** Information on the irrigation scheduling carried out at Mocambinho with the PM equation and original and calibrated empirical equations considering different irrigation intervals (II). All the variables, except for II, are expressed in mm.

Equation	II (d)	ETc	ETc (true)	NID	Pe	Pe (true)	Deficit	Excess
PM	1	1809	1809	1651	158	158	-	-
	2	1768	1768	1550	213	213	-	-
	4	1680	1680	1456	219	219	-	-
	6	1569	1569	1292	271	271	-	-
	8	1423	1423	1157	259	259	-	-
HS	1	2133	1808	1935	199	158	1	285
	2	2078	1768	1815	256	213	0	265
	4	1954	1680	1683	264	219	0	227
	6	1788	1569	1459	320	271	0	166
	8	1571	1423	1259	301	259	0	102

HS_cal	1	1781	1764	1606	175	179	44	21
	2	1742	1726	1508	228	228	43	16
	4	1661	1641	1419	236	230	39	15
	6	1559	1538	1266	287	277	31	12
	8	1428	1405	1147	274	261	18	11
ROM	1	2619	1807	2465	154	163	2	823
	2	2524	1766	2310	209	214	2	764
	4	2277	1678	2050	214	219	2	597
	6	1917	1567	1621	279	271	3	332
	8	1572	1419	1295	257	259	4	142
ROM_cal	1	1765	1744	1609	156	235	65	106
	2	1728	1704	1521	203	270	65	101
	4	1647	1622	1434	209	263	58	91
	6	1548	1519	1288	255	297	50	80
	8	1420	1392	1170	245	273	31	60
JH	1	2580	1809	2388	193	158	0	737
	2	2493	1768	2230	254	213	0	680
	4	2279	1680	2006	264	219	0	550
	6	1953	1569	1605	328	271	0	312
	8	1649	1423	1321	308	259	0	164
JH_cal	1	1780	1766	1625	155	171	43	29
	2	1741	1727	1525	211	223	41	27
	4	1657	1644	1436	216	226	37	24
	6	1553	1542	1279	269	276	27	19
	8	1418	1411	1155	255	260	13	13

NID - total net irrigation depth; Pe - effective rainfall; Deficit - ETc deficit in relation to ETc obtained in the irrigation scheduling performed with the PM equation; Excess - excessive irrigation; ETc (true) and Pe (true) - ETc and Pe recalculated using ETo obtained with the PM equation. All the variables are expressed in mm. HS - Hargreaves-Samani; ROM - Romanenko; JH - Jensen-Haise. “\_cal” indicates the calibrated version of an equation.

Among the non-calibrated equations, only the ROM equation used at Viçosa obtained total net irrigation depth close to that obtained with the PM equation. In all other cases, irrigation was overestimated. Thus, such equations promoted excessive water application, increasing the soil moisture above field capacity, as seen in the “Excess” column of Tables 6



and 7. However, after calibration, all the equations obtained total net irrigation depths close to those obtained when using the PM equation. Such behaviors corroborate the reductions in MBE absolute values observed for the estimation of ETo (Tables 4 and 5).

Although the calibrated equations obtained total net irrigation depths close to those obtained using the PM equation, it does not mean that they had the same performance of the PM equation. It may happen that, over the year, the overestimated irrigations have been compensated for the underestimated irrigations, canceling the errors. Thus, irrigation scheduling must be evaluated considering its dynamics over time.

To analyze the performance of the equations considering their time dynamics, it is possible to evaluate the occurrence of excessive water applications, as well as reductions of ETc under adequate irrigation conditions (i.e., irrigation scheduling using the PM equation) in relation to ETc observed under lower water application (i.e., irrigation scheduling using alternative equations). In this sense, although most of the calibrated equations resulted in total net irrigation depths close to those calculated with the PM equation, there were both irrigation underestimation and overestimation during the period evaluated, as shown in columns “Deficit” and “Excess” in Tables 6 and 7. However, after calibrating the equations, there were, in general, large reductions in the excessive water applications. On the other hand, the calibrated equations promoted certain irrigation deficits, slightly reducing total ETc (true) in relation to that observed when scheduling irrigation with the PM equation. For both study sites, the calibrated HS and JH equations were the best options, promoting low excessive water applications and only small reductions in ETc (true).

When scheduling irrigation using variable irrigation intervals, a critical soil water content is adopted to prevent the crop from suffering water deficit. Thus, it is necessary that the current soil water content is always above or, at most, slightly below the critical minimum limit considered. Thus, alternative models for the estimation of ETo must be able to provide sufficiently reliable ETo estimates to meet the condition described above. The results of the irrigation scheduling simulations with variable irrigation intervals are shown in Table 8.

**Table 8.** Information on the irrigation scheduling carried out at Viçosa and Mocambinho with the PM equation and original and calibrated empirical equations using variable irrigation intervals. All the variables are expressed in mm.

Station	Equation	ETc	ETc (true)	NID	Pe	Pe (true)	Deficit ( $0.5 < f \leq 0.6$ )	Deficit ( $f > 0.6$ )	Excess
---------	----------	-----	---------------	-----	----	--------------	-----------------------------------	--------------------------	--------

	PM	1142	1142	866	276	276	5 (7 days)	0	-
	HS	1580	1162	1252	327	249	0	0	340 (81 days)
	HS_cal	1144	1127	816	327	320	29 (24 days)	0	10 (14 days)
Viçosa	ROM	1106	1122	835	271	347	36 (23 days)	52 (11 days)	61 (33 days)
	ROM_cal	1204	1133	916	288	324	33 (21 days)	18 (4 days)	108 (41 days)
	JH	1535	1162	1224	311	262	0	0	324 (78 days)
	JH_cal	1134	1134	854	280	299	22 (16 days)	5 (1 day)	19 (26 days)
	PM	1713	1713	1448	256	256	37 (31 days)	0	-
	HS	2029	1736	1739	284	233	5 (12 days)	0	242 (107 days)
	HS_cal	1687	1670	1411	267	261	80 (47 days)	107 (20 days)	12 (14 days)
Moc.	ROM	2498	1751	2255	238	233	11 (9 days)	0	742 (127 days)
	ROM_cal	1677	1653	1433	240	298	66 (42 days)	247 (38 days)	91 (42 days)
	JH	2460	1754	2198	254	213	0	0	663 (132 days)
	JH_cal	1686	1672	1431	249	259	64 (46 days)	153 (25 days)	24 (38 days)

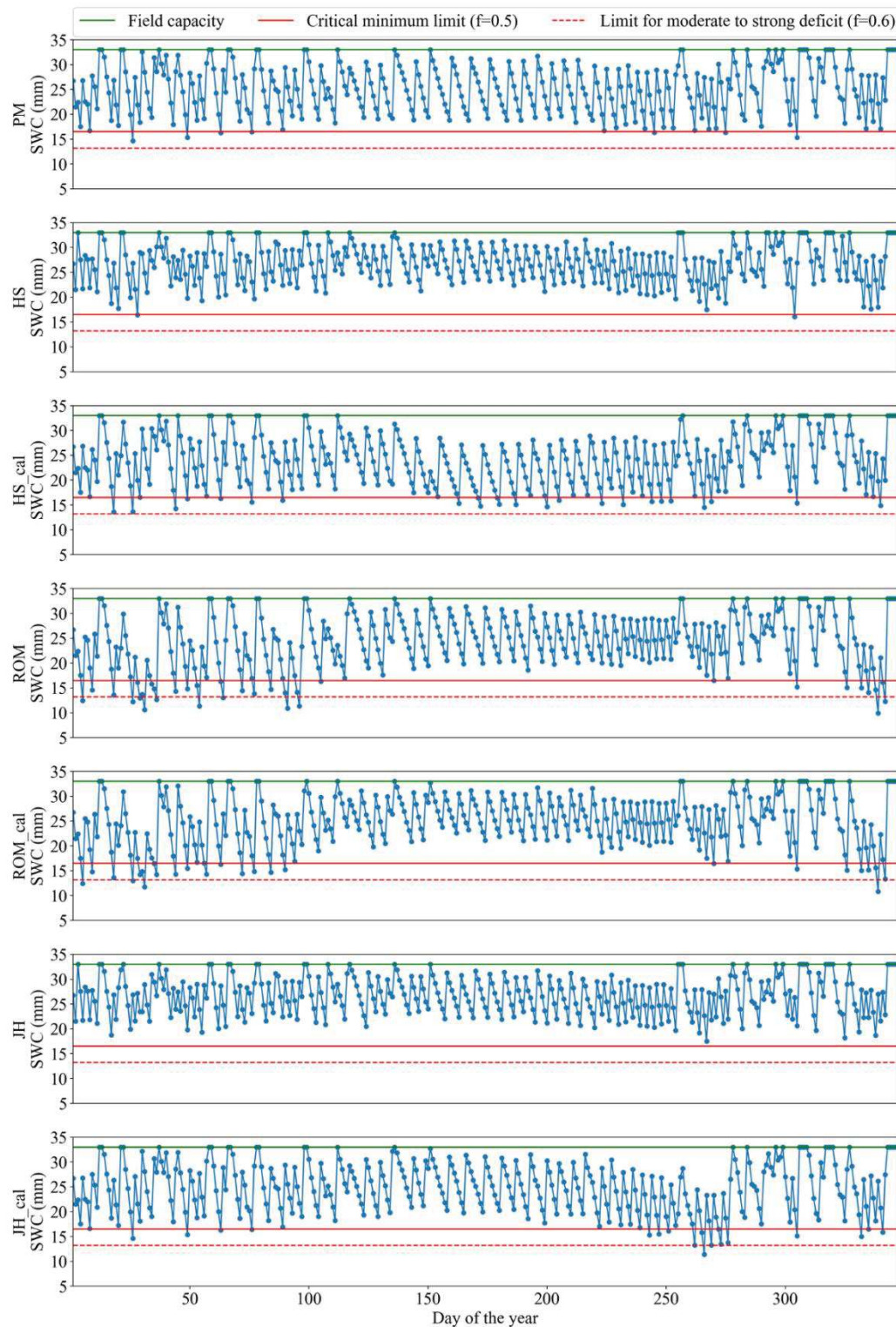
NID - total net irrigation depth; Pe - effective rainfall; Deficit ( $0.5 < f \leq 0.6$ ) and Deficit ( $f > 0.6$ ) - sum of soil water content deficits in relation to the critical level (50% of TAW,  $f = 0.5$ ) in the cases of deficits equivalent to  $0.5 < f \leq 0.6$  and  $f > 0.6$ , respectively (values in parentheses indicate the number of days that the deficits occurred); Excess - excessive irrigation (values in parentheses indicate the number of days that the irrigation excesses occurred); ETc (true) and Pe (true) - ETc and Pe recalculated using ETo obtained with the PM equation. All the variables are expressed in mm. HS - Hargreaves-Samani; ROM - Romanenko; JH - Jensen-Haise. “\_cal” indicates the calibrated version of an equation.

As previously observed, among the non-calibrated equations, only the ROM equation used at Viçosa obtained total net irrigation depth close to that obtained with the PM equation. In the other cases, the total net irrigation depths were much higher than those calculated with the PM equation. After calibrations, there were, in general, reductions in the irrigation excesses.

In relation to the irrigation deficits over the period evaluated, accumulated deficits in relation to the critical soil water content ( $f = 0.5$ ) were computed in two classes: (i) cases in which deficits were equivalent to  $0.5 < f \leq 0.6$  (weak deficit), and (ii)  $f > 0.6$  (moderate to strong deficit). Even using the PM equation, there were some weak deficit events ( $0.5 < f \leq 0.6$ ). This behavior is expected because even though the soil has not reached the limit water content for irrigation (in this study, irrigation was carried out when the soil water content was 40% below TAW) on a particular day, it is possible that, on the next day, the soil water content is already below the critical limit adopted (50% of TAW). However, it is expected that this level of stress,

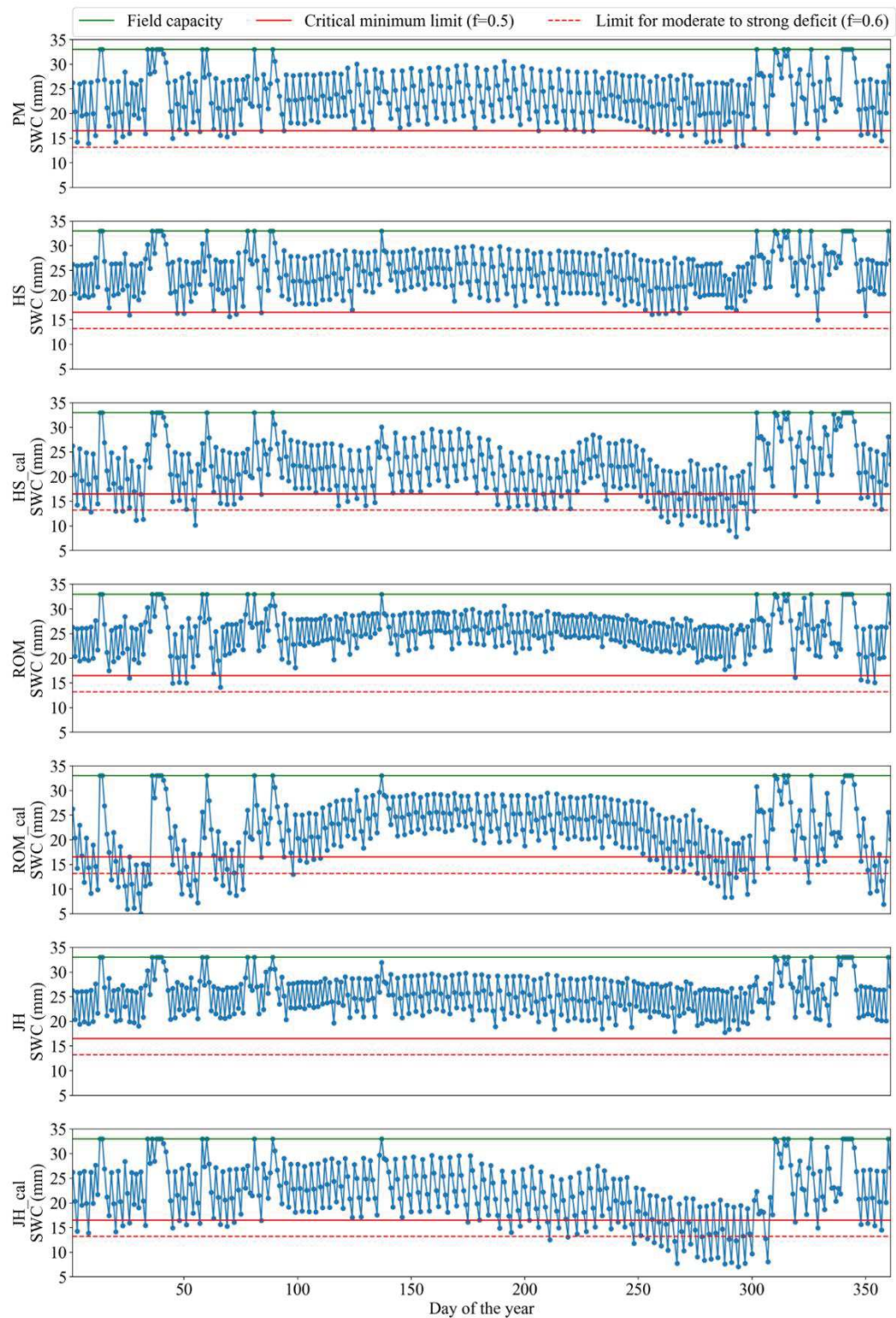
which remains for a short period and is of low intensity, does not cause significant damage to the crops.

At Viçosa, the calibrated HS and JH equations performed the best, with similar performance to each other. For Mocambinho, these equations also obtained the best performances; however, the calibrated HS equation was slightly better than the calibrated JH equation since it had lower moderate to strong deficits ( $f > 0.6$ ) and lower irrigation excesses. These behaviors partially contradict the results obtained when directly evaluating the equations for the estimation of ETo (Tables 4 and 5), since the calibrated JH equation was considered better than the calibrated HS equation in all the studied scenarios. To better assess the irrigation scheduling carried out with the different equations for the estimation of ETo, the soil water content behaviors during the evaluation period at Viçosa and Mocambinho are shown in Figs 2 and 3, respectively. After the end of the irrigation scheduling simulations with each empirical equation, the soil water contents were recalculated based on ETo obtained with the PM equation, as shown in Fig 1. The information presented in Figs 2 and 3 is referring to these recalculated water contents. On the days when there was irrigation, the water contents presented refer to the moment before irrigation.



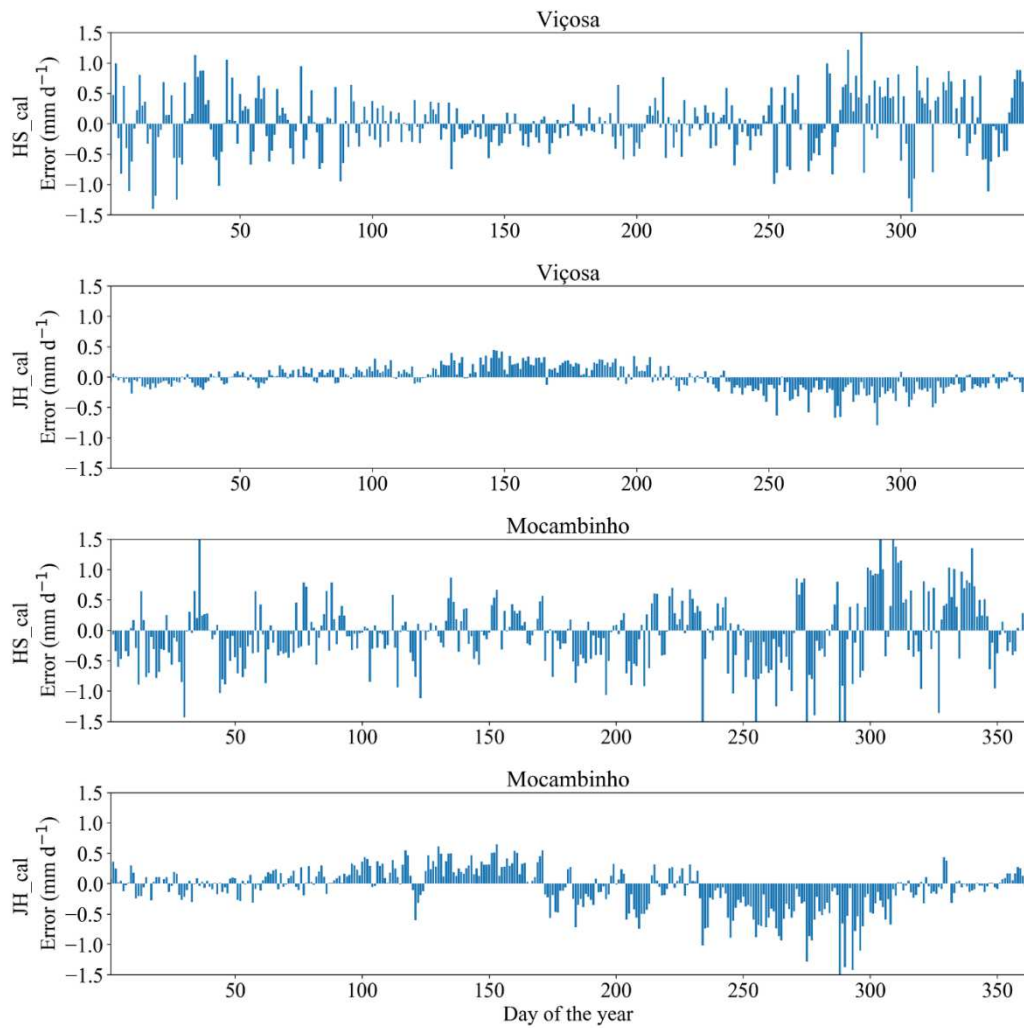
**Fig 2.** Soil water content (SWC) for irrigation scheduling with variable irrigation intervals at Viçosa using the PM equation and the original and calibrated HS, ROM and JH equations. SWC values presented were recalculated at the end of the irrigation scheduling (performed with the empirical equations) using ETo obtained by the PM equation. Days are numbered according to their order throughout the test year.





**Fig 3.** Soil water content (SWC) for irrigation scheduling with variable irrigation intervals at Mocambinho using the PM equation and the original and calibrated HS, ROM and JH equations. SWC values presented were recalculated at the end of the irrigation scheduling (performed with the empirical equations) using ETo obtained by the PM equation. Days are numbered according to their order throughout the test year.

At Viçosa, both the calibrated HS and JH equations promoted only small water deficits below the critical limit (Fig 2). At Mocambinho, when using the calibrated HS and JH equations, the soil water content falls considerably in the period of 250 to 300 days, especially for the calibrated JH equation (Fig 3). Even though the calibrated JH equation presented better metrics than the calibrated HS equation for the estimation of ETo at Mocambinho (Table 5), this equation had continuous ETo underestimations in the period around 250-300 days (Fig 4). On the other hand, the calibrated HS equation, despite showing, in general, greater deviations in relation to ETo obtained with the PM equation, had more alternate ETo underestimates and overestimates, which contributes to partially cancel the errors occurred during the irrigation scheduling period. Similar behavior was observed for Viçosa (Fig 4). It is also worth mentioning that in places with high rainfall levels, problems with ETo underestimation tend to be reduced.



**Fig 4.** Deviation of ETo estimated by the calibrated Hargreaves-Samani (HS) and Jensen-Haise (JH) equations in relation to ETo obtained with the PM equation at Viçosa and Mocambinho. Days are numbered according to their order throughout the test year.

Finally, in addition to evaluating alternative models for the estimation of ETo using error metrics such as RMSE, MAE, MBE and  $R^2$ , it is also important to analyze their behavior throughout the year. Furthermore, the simulation of the use of these models for irrigation scheduling can help in choosing the best model. Future studies could address the development of software, including integration with crop models, for simulating the use of models for the estimation of ETo for irrigation scheduling. Thus, it would be possible to evaluate the performance of the models considering more specific scenarios of interest. It could be considered, for example, the application of an empirical equation only at a certain period of the year, irrigation of crops with different cycles, and different types of soil, among other factors. Another important issue to be considered in future studies and a limitation of the present study



is the use of real field data as benchmark, such as eddy covariance and/or soil water content measurements.

#### **4 Conclusions**

Alternative models for the estimation of ETo are typically assessed using error metrics. However, the model with the best metrics for the estimation of ETo may not be the best option to be used for irrigation scheduling. Despite the importance of the development of methodologies for a better assessment of the performance of models for the estimation of ETo for irrigation scheduling purposes, according to our knowledge, so far, this type of study has not been found. Thus, this study analyzes the performance of three original and calibrated empirical equations, initially assessed using traditional error metrics, for irrigation scheduling, considering the simulation of different irrigation intervals. Two study sites, Viçosa-MG and Mocambinho-MG, Brazil, were used.

In general, the original empirical equations did not perform well for the estimation of ETo, with the exception of the Romanenko and Hargreaves-Samani equations used at Viçosa and Mocambinho, respectively. Calibration promoted performance gains, reducing the tendency of the equations to overestimate ETo. The increase in the time scale also led to reductions in estimation errors.

When used for irrigation scheduling, the calibrated Hargreaves-Samani and Jensen-Haise equations showed the best performances in both Viçosa and Mocambinho stations. Even with greater errors when estimating ETo, the calibrated Hargreaves-Samani equation performed similarly or better than the calibrated Jensen-Haise equation, as it had errors with greater potential to be canceled during the soil water balance. The results obtained are dependent of the climate conditions of the study site, thus, the performance of the equations can be very different in areas with different climatic conditions.

Finally, it is suggested that the assessment of models for the estimation of ETo for use in irrigation scheduling, in addition to using traditional error metrics, consider the performance of the models throughout the year. Furthermore, simulating the application of the models in irrigation scheduling can provide valuable information for choosing the most suitable option.

#### **Acknowledgments**

The authors wish to thank the Brazilian National Institute of Meteorology (INMET) for the meteorological data used.

## References

1. Gu Z, Qi Z, Burghate R, Yuan S, Jiao X, Xu J. Irrigation Scheduling Approaches and Applications: A Review. *J Irrig Drain Eng*. 2020. doi:10.1061/(asce)ir.1943-4774.0001464
2. Pereira LS, Allen RG, Smith M, Raes D. Crop evapotranspiration estimation with FAO56: Past and future. *Agricultural Water Management*. 2015. pp. 4–20. doi:10.1016/j.agwat.2014.07.031
3. Vellidis G, Liakos V, Andreis JH, Perry CD, Porter WM, Barnes EM, et al. Development and assessment of a smartphone application for irrigation scheduling in cotton. *Comput Electron Agric*. 2016;127: 249–259. doi:10.1016/j.compag.2016.06.021
4. Santos IS, Mantovani EC, Venancio LP, da Cunha FF, Aleman CC. Controlled water stress in agricultural crops in brazilian cerrado. *Biosci J*. 2020;36: 886–895. doi:10.14393/BJ-v36n3a2020-47862
5. Ferreira LB, Cunha FF da, Oliveira RA de, Rodrigues TF. A smartphone APP for weather-based irrigation scheduling using artificial neural networks. *Pesqui Agropecuária Bras*. 2020;55. doi:10.1590/s1678-3921.pab2020.v55.01839
6. Pereira LS, Paredes P, Jovanovic N. Soil water balance models for determining crop water and irrigation requirements and irrigation scheduling focusing on the FAO56 method and the dual Kc approach. *Agricultural Water Management*. 2020. doi:10.1016/j.agwat.2020.106357
7. Ramadan KM, Oates MJ, Molina-Martinez JM, Ruiz-Canales A. Design and implementation of a low cost photovoltaic soil moisture monitoring station for irrigation scheduling with different frequency domain analysis probe structures. *Comput Electron Agric*. 2018. doi:10.1016/j.compag.2017.12.038
8. Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56. FAO. Rome, Italy; 1998.
9. Almorox J, Senatore A, Quej VH, Mendicino G. Worldwide assessment of the Penman–Monteith temperature approach for the estimation of monthly reference evapotranspiration. *Theor Appl Climatol*. 2018;131: 693–703. doi:10.1007/s00704-016-1996-2
10. Wu L, Fan J. Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. *PLoS One*. 2019;14: e0217520. doi:10.1371/journal.pone.0217520

11. Antonopoulos VZ, Antonopoulos A V. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Comput Electron Agric.* 2017;132: 86–96. doi:10.1016/j.compag.2016.11.011
12. Ferreira LB, da Cunha FF, de Oliveira RA, Fernandes Filho EI. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. *J Hydrol.* 2019;572: 556–570. doi:10.1016/j.jhydrol.2019.03.028
13. Pandey PK, Dabral PP, Pandey V. Evaluation of reference evapotranspiration methods for the northeastern region of India. *Int Soil Water Conserv Res.* 2016;4: 52–63. doi:10.1016/j.iswcr.2016.02.003
14. Hadria R, Benabdelouhab T, Lionbouli H, Salhi A. Comparative assessment of different reference evapotranspiration models towards a fit calibration for arid and semi-arid areas. *J Arid Environ.* 2021. doi:10.1016/j.jaridenv.2020.104318
15. Quej VH, Almorox J, Arnaldo JA, Moratiel R. Evaluation of temperature-based methods for the estimation of reference evapotranspiration in the Yucatán Peninsula, Mexico. *J Hydrol Eng.* 2019;24: 1–10. doi:10.1061/(ASCE)HE.1943-5584.0001747
16. Liu X, Xu C, Zhong X, Li Y, Yuan X, Cao J. Comparison of 16 models for reference crop evapotranspiration against weighing lysimeter measurement. *Agric Water Manag.* 2017;184: 145–155. doi:10.1016/j.agwat.2017.01.017
17. Bernardo S, Mantovani EC, Silva DD, Soares AA. *Manual de Irrigação*. 9th ed. Viçosa: Editora UFV; 2019.
18. Hargreaves GH, Samani ZA. Reference Crop Evapotranspiration from Temperature. *Appl Eng Agric.* 1985;1: 96–99. doi:10.13031/2013.26773
19. Mehdizadeh S, Behmanesh J, Khalili K. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput Electron Agric.* 2017;139: 103–114. doi:10.1016/j.compag.2017.05.002
20. Jensen M, Haise H. Estimating Evapotranspiration from Solar Radiation. *Proc Am Soc Civ Eng J Irrig Drain Div.* 1963;89: 15–41.
21. Hargreaves GH, Allen RG. History and evaluation of hargreaves evapotranspiration equation. *J Irrig Drain Eng.* 2003;129: 53–63. doi:10.1061/(ASCE)0733-9437(2003)129:1(53)

## General conclusions

The application of machine learning techniques, which include traditional machine learning models and deep learning models, showed great potential for modeling reference evapotranspiration (ET<sub>o</sub>) in the various conditions evaluated.

The use of hourly temperature and relative humidity data, combined with hourly extraterrestrial solar radiation, proved to be a very promising approach to estimate daily ET<sub>o</sub> in scenarios with limited availability of meteorological data. In this approach, 1D convolutional neural networks (1D CNN) provided better results than the other models evaluated.

When predicting ET<sub>o</sub> for the next seven days, in general, the MIMO (multiple input multiple output) prediction strategy was the best alternative, offering good performance and lower computational cost. The evaluated deep learning models performed slightly better than the traditional machine learning models, and both approaches resulted in better performances than those obtained using historical monthly means as a prediction of future ET<sub>o</sub> values.

By employing machine learning models to estimate ET<sub>o</sub> and commonly unavailable meteorological data (relative humidity, solar radiation and wind speed), superior performances were observed in relation to those obtained with traditional methodologies. The use of multi-task learning to estimate, in a combined way, missing meteorological data and ET<sub>o</sub> resulted in performances similar to those found when considering individual estimates (single-task learning).

Finally, it was found that in addition to using error metrics, such as root mean square error (RMSE), the evaluation of ET<sub>o</sub> models should also consider the behavior of the models throughout the year. Furthermore, simulating the application of ET<sub>o</sub> models in irrigation scheduling can provide valuable information for choosing the most appropriate model.