



# Multivariate adaptive regression splines (MARS) applied to daily reference evapotranspiration modeling with limited weather data

Lucas Borges Ferreira<sup>1\*</sup>, Anunciene Barbosa Duarte<sup>2</sup> , Fernando França da Cunha<sup>1</sup> and Elpídio Inácio Fernandes Filho<sup>3</sup>

<sup>1</sup>Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Av PH Rolfs, s/n., Campos Universitário, 36570-900, Viçosa, Minas Gerais, Brazil.

<sup>2</sup>Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil. <sup>3</sup>Departamento de Solos, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil. \*Author for correspondence. E-mail: contato.lucasbf@gmail.com

**ABSTRACT.** Estimation of reference evapotranspiration ( $ET_0$ ) is very relevant for water resource management. The Penman-Monteith (PM) equation was proposed by the Food and Agriculture Organization (FAO) as the standard method for estimation of  $ET_0$ . However, this method requires various weather data, such as air temperature, wind speed, solar radiation and relative humidity, which are often unavailable. Thus, the objective of this study was to compare the performance of multivariate adaptive regression splines (MARS) and alternative equations, in their original and calibrated forms, to estimate daily  $ET_0$  with limited weather data. Daily data from 2002 to 2016 from 8 Brazilian weather stations were used.  $ET_0$  was estimated using empirical equations, PM equation with missing data and MARS. Four data availability scenarios were evaluated as follows: temperature only, temperature and solar radiation, temperature and relative humidity, and temperature and wind speed. The MARS models demonstrated superior performance in all scenarios. The models that used solar radiation showed the best performance, followed by those that used relative humidity and, finally, wind speed. The models based only on air temperature had the worst performance.

**Keywords:** data driven; irrigation scheduling; agrometeorology; artificial intelligence.

Received on October 2, 2017.

Accepted on January 9, 2018.

## Introduction

Evapotranspiration is one of the main components of the water cycle, allowing the transfer of water and energy into the atmosphere (Fernandes, Paiva, & Rotunno Filho, 2012). Its estimation is very relevant for decision making regarding water use, irrigation scheduling, environmental studies, and others (Pereira, Allen, Smith, & Raes, 2015).

The Food and Agriculture Organization (FAO) proposed the Penman-Monteith (PM) equation as a standard method for estimation of reference evapotranspiration ( $ET_0$ ) (Allen, Pereira, Raes, & Smith, 1998). It is an equation that, due to its physical basis, requires several climatic parameters, such as air temperature, relative humidity, wind speed and solar radiation. The large number of required climatic variables is one of the factors responsible for the satisfactory performance of this method; however, its use is limited, since these data are commonly unavailable or unreliable in several regions of the world (Talaee, 2014), especially in developing countries (Djaman, Irmak, & Futakuchi, 2017), such as Brazil.

To estimate  $ET_0$  with limited weather data, many studies have been conducted by using a reduced number of variables and developing empirical and semi-empirical models based on temperature (Hargreaves & Samani, 1985; Oudin et al., 2005), temperature and solar radiation (Makkink, 1957; Jensen & Haise, 1963), temperature and relative humidity (Valiantzas, 2013), and others. These methods, unlike the PM equation, which can be used globally without additional adjustments (Pereira et al., 2015), require local calibrations to obtain more satisfactory performances (Gao, Peng, Xu, Yang, & Wang, 2015).

In addition to conventional equations, such as those mentioned above, data driven methods have been widely used in recent years. Among these, artificial neural network (ANN) (Yassin, Alazba, & Mattar, 2016; Antonopoulos & Antonopoulos, 2017), support vector machine (SVM) (Tabari, Kisi, Ezani, & Talaee, 2012; Mehdizadeh, Behmanesh, & Khalili, 2017), extreme learning machine (ELM) (Feng, Cui, Zhao, Hu, & Gong, 2016;

Gocic, Petković, Shamshirband, & Kamsin, 2016) and adaptive neuro-fuzzy inference system (ANFIS) (Tabari et al., 2012) can be cited. These methods present, in general, higher performances compared to the conventional methods for estimation of  $ET_0$  and, according to Mehdizadeh et al. (2017), can be applied when modeling complex non-linear problems. Thus, data driven models are powerful tools for  $ET_0$  modeling since, according to Feng et al. (2016), it is a complex and non-linear task.

Among data driven methods, multivariate adaptive regression splines (MARS) is a promising technique for estimation of  $ET_0$ ; however, for this purpose, it has been poorly explored. MARS is a type of regression proposed by Friedman (1991) that is capable of modeling complex relations between a response variable and a set of predictor variables. According to Koc and Bozdogan (2015), this technique has been applied successfully in several areas of knowledge, such as medicine, business, molecular biology and several others. In addition to its potential for modeling, MARS also has the able advantage of being used in the form of an explicit algebraic equation, unlike other data driven methods, such as ANN, SVM, ELM, and ANFIS.

In this context, this study aims to compare the performance of MARS and alternative equations (in their original and calibrated forms) to estimate daily  $ET_0$  with limited weather data.

## Material and methods

### Database and study sites

To carry out this study, daily data from 2002 to 2016 were collected from 8 weather stations of the National Institute of Meteorology (INMET) of Brazil, available in the Meteorological Database for Teaching and Research (BDMEP) (Table 1). The stations were selected to cover various climatic conditions and, consequently, to make the results more representative. Thus, stations were selected in several regions of Brazil, and their main characteristics are presented in Table 2.

**Table 1.** Location, altitude and climate classification of the chosen weather stations.

Station	Latitude (°)	Longitude (°)	Altitude (m)	Köppen classification
Araxá - MG	-19.60	-46.94	1024	Cwb
Cabrobó - PE	-8.51	-39.33	341	Bsh
Curitiba - PR	-25.43	-49.26	924	Cfb
Eirunepé - AM	-6.66	-69.86	104	Af
Lages - SC	-27.81	-50.33	937	Cfb
Macapá - AP	-0.05	-51.11	14	Am
Palmas - TO	-10.19	-48.30	280	Aw
Santa Maria - RS	-29.70	-53.70	95	Cfa

**Table 2.** Daily mean values and standard deviations of meteorological variables for the chosen stations.

Station	Tmax	Tmin	n	RH	U <sub>2</sub>	ET <sub>0</sub>
Araxá - MG	27.8 (±2.8)	17.0 (±2.5)	6.9 (±3.4)	67.8 (±15.2)	1.9 (±0.7)	3.9 (±1.1)
Cabrobó - PE	32.6 (±2.5)	22.2 (±1.7)	8.0 (±3.0)	59.8 (±13.9)	3.1 (±1.2)	5.9 (±1.7)
Curitiba - PR	24.0 (±4.7)	13.7 (±4.0)	5.1 (±3.6)	80.3 (±9.4)	1.6 (±0.6)	2.9 (±1.3)
Eirunepé - AM	33.5 (±2.5)	21.7 (±1.7)	3.9 (±2.9)	86.0 (±5.1)	0.1 (±0.2)	3.2 (±0.8)
Lages - SC	22.3 (±5.2)	11.9 (±4.8)	5.4 (±3.7)	81.3 (±9.3)	0.9 (±0.6)	2.7 (±1.3)
Macapá - AP	32.1 (±1.8)	24.0 (±0.8)	6.8 (±3.2)	80.5 (±7.9)	1.5 (±0.8)	4.3 (±1.2)
Palmas - TO	33.8 (±2.8)	22.3 (±1.8)	6.9 (±3.3)	66.6 (±17.1)	0.9 (±0.7)	4.3 (±1.2)
Santa Maria - RS	25.5 (±6.1)	14.7 (±5.5)	6.2 (±4.1)	77.7 (±11.6)	1.5 (±0.7)	3.1 (±1.7)

Tmax - maximum temperature; Tmin - minimum temperature; n - sunshine duration; RH - relative humidity; U<sub>2</sub> - wind speed at 2 m height, ET<sub>0</sub> - reference evapotranspiration estimated by the Penman-Monteith equation.

Maximum and minimum air temperature (°C), relative humidity (%), sunshine duration (h) and wind speed at 10 m height (m s<sup>-1</sup>) were obtained. The wind speed was converted to 2 m height, and the solar radiation was estimated based on sunshine duration, as recommended by Allen et al. (1998).

The collected data were submitted to a preprocessing, which eliminated days with missing data or the presence of a minimum temperature that was higher than the maximum temperature, negative sunshine duration or sunshine duration higher than the photoperiod, negative or greater than 100% relative humidity and wind speed (at 10 m height) with negative or greater than 20 m s<sup>-1</sup> value.

### Methods for estimation of ET<sub>o</sub>

The Penman-Monteith (PM) equation (Equation 1), using all required measured meteorological data, was used as a reference to estimate daily ET<sub>o</sub>.

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

where: ET<sub>oPM</sub> represents the reference evapotranspiration estimated by the Penman-Monteith equation (mm day<sup>-1</sup>), R<sub>n</sub> represents the net solar radiation (MJ m<sup>-2</sup> day<sup>-1</sup>), G represents the soil heat flux (MJ m<sup>-2</sup> day<sup>-1</sup>; considered to be null for daily estimates), T represents daily mean air temperature (°C), U<sub>2</sub> represents the wind speed at a 2 m height (m s<sup>-1</sup>), e<sub>s</sub> represents the saturation vapor pressure (kPa), e<sub>a</sub> represents the actual vapor pressure (kPa), Δ represents the slope of the saturation vapor pressure function (kPa °C<sup>-1</sup>), and γ represents the psychrometric constant (kPa °C<sup>-1</sup>).

To evaluate the performance of MARS against conventional equations, ET<sub>o</sub> was estimated in four measured data availability scenarios: temperature data only, temperature and solar radiation, temperature and relative humidity, and temperature and wind speed. To accomplish this, besides MARS, empirical equations and the Penman-Monteith equation with missing data were used.

The empirical equations used were the Hargreaves-Samani (Hargreaves & Samani, 1985), Oudin (Oudin et al., 2005), Makkink (Makkink, 1957), Jensen and Haise (Jensen & Haise, 1963), Romanenko (presented in Mehdizadeh et al., 2017) and Valiantzas (Valiantzas, 2013) equations, which are presented in Equations 2, 3, 4, 5, 6, and 7, respectively.

$$ET_o = 0.0023 \frac{R_a}{\lambda} (T + 17.8) (T_{max} - T_{min})^{0.5} \quad (2)$$

$$ET_o = \frac{R_a T + 5}{\lambda 100} \text{ if } T > -5; 0 \text{ otherwise} \quad (3)$$

$$ET_o = 0.61 \frac{\Delta}{\Delta + \gamma} \frac{R_s}{\lambda} - 0.12 \quad (4)$$

$$ET_o = 0.408 R_s (0.0252 T + 0.078) \quad (5)$$

$$ET_o = 0.00006 (25 + T)^2 (100 - RH) \quad (6)$$

$$ET_o = 0.00668 R_a \sqrt{(T + 9.5)(T_{max} - T_{dew})} - 0.0696 (T_{max} - T_{dew}) - 0.024 (T + 20) \left(1 - \frac{RH}{100}\right) - 0.00455 R_a \sqrt{(T_{max} - T_{dew})} + 0.0984 (T + 17) (1.03 + 0.00055 TR^2 - \frac{RH}{100}) \quad (7)$$

$$T_{dew} = \frac{116.91 + 237.3 \ln(e_a)}{16.78 - \ln(e_a)} \quad (8)$$

where: ET<sub>o</sub> represents reference evapotranspiration (mm day<sup>-1</sup>), R<sub>a</sub> represents extraterrestrial radiation (MJ m<sup>-2</sup> day<sup>-1</sup>), λ represents latent heat of vaporization (MJ kg<sup>-1</sup>; λ = 2.45 MJ kg<sup>-1</sup> at 20°C), T<sub>max</sub> represents the maximum air temperature (°C), T<sub>min</sub> represents the minimum air temperature (°C), T represents the mean air temperature (°C), Δ represents the slope of the saturation vapor pressure function (kPa °C<sup>-1</sup>), γ represents the psychrometric constant (kPa °C<sup>-1</sup>), RH represents the relative humidity (%), T<sub>dew</sub> represents the dewpoint temperature (°C), TR represents the temperature range (°C), and e<sub>a</sub> represents the actual vapor pressure (kPa).

To estimate ET<sub>o</sub> using the PM equation with missing data, actual vapor pressure and solar radiation were estimated with Equations 9 and 10, respectively, and wind speed was set at 2 m s<sup>-1</sup>, as recommended by Allen et al. (1998).

$$e_a = 0.611 \exp \left[ \frac{17.27 T_{min}}{T_{min} + 237.3} \right] \quad (9)$$

where: e<sub>a</sub> represents the actual vapor pressure (kPa), and T<sub>min</sub> represents the minimum air temperature (°C).

$$R_s = 0.16 R_a (T_{max} - T_{min})^{0.5} \quad (10)$$

where: R<sub>s</sub> represents solar radiation (MJ m<sup>-2</sup> day<sup>-1</sup>), R<sub>a</sub> represents extraterrestrial radiation (MJ m<sup>-2</sup> day<sup>-1</sup>), T<sub>max</sub> represents the maximum air temperature (°C), and T<sub>min</sub> represents the minimum air temperature (°C).

To improve the performance of the equations studied, a local calibration was performed with a simple linear regression using data from 2002 to 2011 (10 years), as suggested by Allen et al. (1998). To accomplish this, a

linear regression was fitted so that the  $ET_o$  values estimated by the reference method (i.e., PM with full data set) were set as the dependent variable, and those estimated by the equation to be calibrated were set as the independent variable, according to Equation 11. The obtained intercept (a) and slope (b) were used as local calibration parameters.

$$ET_{ocal} = a + b(ET_o) \quad (11)$$

where:  $ET_{ocal}$  represents the reference evapotranspiration estimated by the calibrated equation ( $\text{mm day}^{-1}$ ), a and b represent the calibration parameters, and  $ET_o$  represents the reference evapotranspiration estimated by the equation to be calibrated ( $\text{mm day}^{-1}$ ).

The MARS method is a nonparametric multivariate regression technique that can map relations between input and output variables without assumptions, model nonlinearities and interactions, and automatically choose the variables that are important for the modeling process. In MARS, base functions are fitted at different intervals of the independent variables. The initial and final points of these intervals are called knots (Mehdizadeh et al., 2017). A MARS model with 2 base functions can be seen in Figure 1.

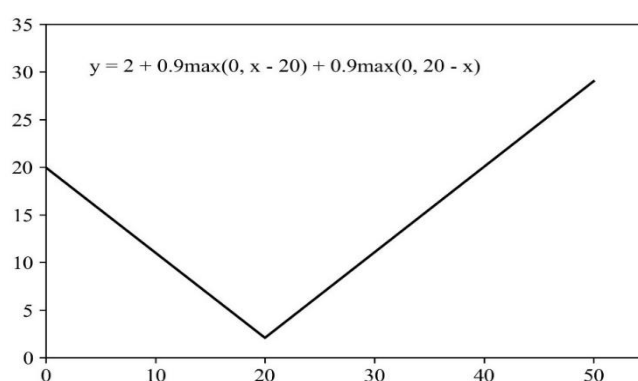


Figure 1. MARS model with two base functions.

To evaluate the performance of the MARS models and the equations studied, these were divided according to the weather data required for each model (Table 3).

Table 3. Methods used in the study and their respective input variables.

Model basis	Model	Tmax	Tmin	Ra	Rs	RH	$U_2$
Temperature	MARS1	✓	✓	✓			
	PMT	✓	✓	✓			
	HS	✓	✓	✓			
	ODU	✓	✓	✓			
Temperature and solar radiation	MARS2	✓	✓		✓		
	PMR	✓	✓	✓	✓		
	MAK	✓	✓		✓		
	JH	✓	✓		✓		
Temperature and relative humidity	MARS3	✓	✓	✓		✓	
	PMH	✓	✓	✓		✓	
	ROM	✓	✓			✓	
	VLT	✓	✓	✓		✓	
Temperature and wind speed	MARS4	✓	✓	✓			✓
	PMW	✓	✓	✓			✓

Tmax - maximum air temperature; Tmin - minimum air temperature; Ra - extraterrestrial radiation; Rs - solar radiation; RH - relative humidity;  $U_2$  - wind speed at 2 m height; MARS - multivariate adaptive regression splines; PMT - Penman-Monteith with measured data of temperature; HS - Hargreaves-Samani; OUD - Oudin; PMR - Penman-Monteith with measured data of temperature and solar radiation; MAK - Makink; JH - Jensen-Haise; PMH - Penman-Monteith with measured data of temperature and relative humidity; ROM - Romanenko; VLT - Valiantzas; PMW - Penman-Monteith with measured data of temperature and wind speed.

The development process of a MARS model involves two steps; in the first step (i.e., forward step), an over-fitted model is produced with a large number of knots. In the second step (i.e., backward step), a pruning technique is applied to remove redundant knots (Kisi, 2015). More details about MARS can be found in Cheng and Cao (2014).

MARS models were developed using  $ET_o$  estimated by the PM equation (with full data set) as benchmark, with data from 2002 to 2011 (10 years). The implementation process was performed using the py-earth library for the Python programming language.

### Performance evaluation

The performance of the models was evaluated using data from 2012 to 2016 (5 years). To accomplish this, the statistical indices root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination ( $R^2$ ) were used based on the following equations.

$$RMSE = \sqrt{\frac{1}{n} \sum (P_i - O_i)^2} \quad (12)$$

$$MAE = \frac{1}{n} \sum |P_i - O_i| \quad (13)$$

$$R^2 = \left( \frac{\sum (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{(\sum (P_i - \bar{P})^2)(\sum (O_i - \bar{O})^2)}} \right)^2 \quad (14)$$

where: RMSE is the root mean square error ( $\text{mm day}^{-1}$ ), MAE is the mean absolute error ( $\text{mm day}^{-1}$ ),  $R^2$  is the coefficient of determination,  $P_i$  represents the predicted value ( $\text{mm day}^{-1}$ ),  $O_i$  represents the observed value ( $\text{mm day}^{-1}$ ),  $\bar{P}$  represents the mean of predicted values ( $\text{mm day}^{-1}$ ),  $\bar{O}$  represents the mean of observed values ( $\text{mm day}^{-1}$ ), and  $n$  represents the number of data pairs.

## Results and discussion

### Temperature-based models

By evaluating the performance of the PMT, HS and OUD methods, it was observed that, in general, the PMT method had the best performance, with a lower RMSE and MAE values, while the OUD method had the worst performance (Table 4). These results corroborate with Almorox, Senatore, Quej, and Mendicino (2018), who concluded that the PMT equation presents  $ET_o$  estimates at a monthly scale more accurate compared to the HS equation in different regions of the world. Similarly, Alencar, Sediya, and Mantovani (2015) obtained better performances of the PMT equation compared to the HS equation on a daily scale in a study carried out in Brazil. The better performance of the HS equation compared to the OUD equation was also reported by Almorox, Quej, and Martí (2015).

**Table 4.** Statistical indices for the temperature-based models.

Station	Index	MARS1	PMT	PMTcal	HS	HScal	OUD	OUDcal
Araxá	RMSE	0.66	0.70	0.70	0.79	0.74	1.04	1.02
	MAE	0.53	0.56	0.56	0.63	0.59	0.86	0.82
	$R^2$	0.71	0.68	0.68	0.65	0.65	0.33	0.33
Cabrobó	RMSE	0.88	1.90	0.92	1.57	0.94	1.85	1.20
	MAE	0.70	1.69	0.73	1.37	0.76	1.62	0.98
	$R^2$	0.68	0.66	0.66	0.65	0.65	0.43	0.43
Curitiba	RMSE	0.53	0.67	0.55	0.86	0.56	0.96	0.84
	MAE	0.39	0.55	0.41	0.71	0.42	0.76	0.70
	$R^2$	0.84	0.83	0.83	0.83	0.83	0.59	0.59
Eirunepé	RMSE	0.51	1.76	0.54	2.12	0.56	1.71	0.73
	MAE	0.39	1.66	0.43	2.03	0.45	1.54	0.62
	$R^2$	0.61	0.56	0.56	0.53	0.53	0.21	0.21
Lages	RMSE	0.52	0.89	0.52	1.10	0.53	0.89	0.76
	MAE	0.39	0.75	0.40	0.94	0.39	0.68	0.61
	$R^2$	0.84	0.84	0.84	0.84	0.84	0.64	0.64
Macapá	RMSE	0.62	0.90	0.77	0.82	0.78	1.12	0.90
	MAE	0.47	0.75	0.61	0.68	0.61	0.88	0.76
	$R^2$	0.71	0.55	0.55	0.55	0.55	0.39	0.39
Palmas	RMSE	0.87	1.10	1.00	1.32	1.05	1.41	1.34
	MAE	0.65	0.91	0.72	1.16	0.76	1.12	1.02
	$R^2$	0.61	0.47	0.47	0.41	0.41	0.05	0.05
Santa Maria	RMSE	0.64	0.85	0.75	1.01	0.73	1.05	1.00
	MAE	0.45	0.67	0.54	0.81	0.51	0.79	0.77
	$R^2$	0.86	0.81	0.81	0.82	0.82	0.67	0.67
Mean	RMSE	0.65	1.10	0.72	1.20	0.74	1.25	0.97
	MAE	0.50	0.94	0.55	1.04	0.56	1.03	0.78
	$R^2$	0.73	0.67	0.67	0.66	0.66	0.41	0.41

MARS1 - multivariate adaptive regression splines with measured data of temperature; PMT - Penman-Monteith with measured data of temperature; HS - Hargreaves-Samani; OUD - Oudin. The expression "cal" indicates the locally calibrated version of a given model.

After local calibration, better performances were obtained for all methods at all sites. The calibrated PMT and HS methods had performances very similar to each other, surpassing the calibrated OUD method. The fact that the OUD method had the worst performance can be justified by the structure of the equation, which was not able to satisfactorily explain the relationship between the input variables (i.e., air temperature and extraterrestrial radiation) and  $ET_o$ , with smaller  $R^2$  values. It is important to note that calibration by simple linear regression does not change the  $R^2$  value. According to Liu et al. (2017), an equation can improve its performance with a local calibration; however, when it presents a failed structure, the structure optimization should receive special attention.

The MARS models that were developed with only measured data of for air temperature (MARS1) had performances significantly superior to the temperature-based equations in their original forms. Considering the calibrated equations, the MARS1 models continued to have superior performances, but with smaller differences. The largest performance improvements were observed at the Macapá and Palmas stations, where the MARS1 models were able to better correlate air temperature and extraterrestrial radiation with  $ET_o$ , which had higher  $R^2$  values and lower RMSE and MAE values. It is important to emphasize that, even though these models had the best performance, the results obtained by the MARS1 models were considered reasonable, as they did not attain a high enough performance.

In general, the obtained results show the complexity involved in the  $ET_o$  modeling process using only measured data of air temperature; even with the best model (MARS1), it was not possible to obtain a high performance. According to Almorox et al. (2015), temperature-based models have low correlations with the PM method (with full data set) in tropical climates, where the role of other climatic variables, such as vapor pressure deficit, can be decisive. Corroborating these authors, it was verified that the highest  $R^2$  values were obtained at the Curitiba, Lages and Santa Maria stations, which are all located in southern Brazil and belong to climatic class C (temperate climate) (Table 1).

### Temperature and solar radiation-based models

As observed in the temperature-based models, the Penman-Monteith equation, when using measured data of temperature and solar radiation (PMR), had a stronger performance when compared to the Makkink (MAK) and Jensen-Haise (JH) equations, with lower RMSE and MAE values (Table 5). The PMR equation was only surpassed at the Cabrobó station, where the JH equation had a better performance, and at the Eirunepé and Lages stations, where the Makkink equation had a best performance. Sentelhas, Gillespie, and Santos (2010) also obtained good results using the PMR equation. On the other hand, the JH equation had the worst performance, corroborating with Cunha, Magalhães, and Castro (2013), which reported the unsatisfactory performance of the JH equation in its original form.

After calibration, all of the equations obtained more accurate results; there were close performances among them, with slight superiority for the JH equation. It is important to highlight that the results obtained by all of the calibrated equations were quite satisfactory and had a strong agreement with  $ET_o$  estimated by the PM equation with full data set. Lower performances were obtained only at the Palmas and Cabrobó stations, but these were still considered reasonable. It should be noted that at the Eirunepé station, the MAK and JS models estimated  $ET_o$  with extremely high precision, obtaining  $R^2$  values equal to 0.98 and 0.99, respectively. Before the calibration the JH equation had the worst performance; however, due to its better structure, evidenced by its higher  $R^2$  values, the local calibration was able to make it quite accurate and surpass the others.

The lower performance observed at the Palmas and Cabrobó stations was possibly related to the higher standard deviation values observed for relative humidity (at Palmas station) and wind speed (at Cabrobó station) (Table 2). The greater oscillation of variables that were not used as inputs may lead to a lower performance of the methods, since these variables have a greater influence on  $ET_o$ , which makes the modeling process even more complex.

In turn, the MARS model developed with temperature and solar radiation data (MARS2) also had a high performance, with a better performance than the non-calibrated equations and a similar performance, that was slightly higher, than the calibrated equations.

### Temperature and relative humidity-based models

By evaluating the performance of the temperature and relative humidity-based models, it was possible to observe that the Romanenko (ROM) equation had the worst performance compared to the Valiantzas (VLT) and Penman-Monteith equations using only measured data of temperature and relative humidity (PMH). The ROM equation, apart from having the highest RMSE and MAE values, also had the lowest  $R^2$  values

(Table 5). The PMH equation performed better than the VLT equation at almost all evaluated sites. In studies performed by Tabari, Grismer and Trajkovic (2013) and Mehdizadeh et al. (2017) in arid and semi-arid regions, the ROM equation had a performance lower than the temperature-based methods, which was also seen in this study.

**Table 5.** Statistical indices for the temperature and solar radiation-based models (a) and statistical indices for the temperature and relative humidity-based models (b).

Station (a)	Index	MARS2	PMR	PMRcal	MAK	MAKcal	JH	JHcal
Araxá	RMSE	0.37	0.41	0.41	0.84	0.42	1.17	0.40
	MAE	0.28	0.33	0.32	0.71	0.30	1.02	0.30
	R <sup>2</sup>	0.91	0.89	0.89	0.89	0.89	0.90	0.90
Cabrobó	RMSE	0.69	1.38	0.69	2.35	0.72	1.04	0.70
	MAE	0.55	1.20	0.55	2.22	0.56	0.84	0.55
	R <sup>2</sup>	0.81	0.82	0.82	0.80	0.80	0.81	0.81
Curitiba	RMSE	0.24	0.41	0.30	0.51	0.34	0.93	0.25
	MAE	0.17	0.35	0.21	0.41	0.25	0.74	0.18
	R <sup>2</sup>	0.97	0.95	0.95	0.93	0.93	0.96	0.96
Eirunepé	RMSE	0.08	1.05	0.22	0.47	0.11	1.72	0.08
	MAE	0.06	1.01	0.17	0.46	0.08	1.61	0.06
	R <sup>2</sup>	0.99	0.92	0.92	0.98	0.98	0.99	0.99
Lages	RMSE	0.20	0.58	0.25	0.33	0.28	0.86	0.22
	MAE	0.15	0.53	0.19	0.27	0.22	0.67	0.17
	R <sup>2</sup>	0.98	0.96	0.96	0.95	0.95	0.97	0.97
Macapá	RMSE	0.25	0.36	0.29	0.82	0.29	2.07	0.27
	MAE	0.18	0.31	0.23	0.75	0.23	1.99	0.21
	R <sup>2</sup>	0.95	0.93	0.93	0.94	0.94	0.95	0.95
Palmas	RMSE	0.73	0.87	0.78	1.27	0.81	1.92	0.76
	MAE	0.44	0.71	0.50	0.97	0.49	1.75	0.46
	R <sup>2</sup>	0.72	0.67	0.67	0.67	0.67	0.70	0.70
Santa Maria	RMSE	0.40	0.63	0.50	0.71	0.54	1.33	0.43
	MAE	0.26	0.50	0.33	0.51	0.36	1.07	0.28
	R <sup>2</sup>	0.95	0.91	0.91	0.90	0.90	0.94	0.94
Mean	RMSE	0.37	0.71	0.43	0.91	0.44	1.38	0.39
	MAE	0.26	0.62	0.31	0.79	0.31	1.21	0.28
	R <sup>2</sup>	0.91	0.88	0.88	0.88	0.88	0.90	0.90
Station (b)	Index	MARS3	PMH	PMHcal	ROM	ROMcal	VLT	VLTcal
Araxá	RMSE	0.46	0.51	0.51	1.79	0.84	0.75	0.50
	MAE	0.35	0.41	0.40	1.37	0.70	0.60	0.40
	R <sup>2</sup>	0.86	0.83	0.83	0.54	0.54	0.83	0.83
Cabrobó	RMSE	0.71	1.29	0.76	2.36	0.94	1.98	0.75
	MAE	0.55	1.10	0.60	1.99	0.76	1.78	0.59
	R <sup>2</sup>	0.80	0.76	0.76	0.72	0.72	0.77	0.77
Curitiba	RMSE	0.38	0.46	0.45	0.91	0.72	0.57	0.45
	MAE	0.29	0.35	0.34	0.76	0.58	0.41	0.34
	R <sup>2</sup>	0.92	0.89	0.89	0.70	0.70	0.89	0.89
Eirunepé	RMSE	0.49	1.09	0.52	1.21	0.64	0.58	0.52
	MAE	0.38	0.96	0.42	1.01	0.50	0.47	0.41
	R <sup>2</sup>	0.64	0.59	0.59	0.39	0.39	0.59	0.59
Lages	RMSE	0.37	0.50	0.43	1.14	0.75	0.48	0.42
	MAE	0.28	0.36	0.32	0.95	0.59	0.35	0.32
	R <sup>2</sup>	0.92	0.88	0.88	0.70	0.70	0.89	0.89
Macapá	RMSE	0.46	0.78	0.54	0.94	0.51	1.33	0.55
	MAE	0.34	0.66	0.41	0.78	0.39	1.16	0.42
	R <sup>2</sup>	0.85	0.79	0.79	0.83	0.83	0.78	0.78
Palmas	RMSE	0.79	0.95	0.82	2.89	0.87	0.87	0.82
	MAE	0.56	0.76	0.58	2.13	0.68	0.62	0.59
	R <sup>2</sup>	0.68	0.64	0.64	0.61	0.61	0.64	0.64
Santa Maria	RMSE	0.49	0.54	0.54	0.99	0.86	0.73	0.56
	MAE	0.37	0.38	0.37	0.78	0.68	0.52	0.40
	R <sup>2</sup>	0.92	0.90	0.90	0.76	0.76	0.89	0.89
Mean	RMSE	0.52	0.76	0.57	1.53	0.77	0.91	0.57
	MAE	0.39	0.62	0.43	1.22	0.61	0.74	0.43
	R <sup>2</sup>	0.82	0.78	0.78	0.65	0.65	0.79	0.79

MARS2 - multivariate adaptive regression splines with measured data of temperature and solar radiation; PMR - Penman-Monteith with measured data of temperature and solar radiation; MAK - Makkink; JH - Jensen-Haise. The expression "cal" indicates the locally calibrated version of a given model.

MARS3 - multivariate adaptive regression splines with measured data of temperature and relative humidity; PMH - Penman-Monteith with measured data of temperature and relative humidity; ROM - Romanenko; VLT - Valiantzas. The expression "cal" indicates the locally calibrated version of a given model.

After calibration, all of the equations obtained a performance increment, where the ROM equation exhibited the worse performance. The PMH and VLT equations had almost identical performances after calibration, which can be explained by the fact that the VLT equation was developed as a simplification of the Penman-Monteith equation (Valiantzas, 2013).

The MARS models developed with measured data of temperature and relative humidity (MARS3) had performances superior to the studied equations, even after calibration. The MARS3 models presented variable performance improvements between stations, with a more significant improvement at the Macapá station, where the RMSE and MAE values reduced from 0.54 and 0.41 to 0.46 and 0.34, respectively, and the  $R^2$  value increased from 0.79 to 0.85 (the PMH equation was the reference equation).

### Temperature and wind speed-based models

The Penman-Monteith equation using only measured data of temperature and wind speed (PMW) produced relatively good results mainly after calibration, with a reduction in RMSE and MAE values (Table 6). For the calibrated equation, RMSE and MAE values ranged from 0.51 to 0.69 and 0.39 to 0.56, respectively, and the  $R^2$  values ranged from 0.45 to 0.86.

**Table 6.** Statistical indices for the temperature and wind speed-based models.

Station	Index	MARS4	PMW	PMWcal
Araxá	RMSE	0.55	0.65	0.64
	MAE	0.42	0.54	0.53
	$R^2$	0.80	0.74	0.74
Cabrobó	RMSE	0.60	1.53	0.69
	MAE	0.50	1.36	0.56
	$R^2$	0.87	0.83	0.83
Curitiba	RMSE	0.47	0.59	0.51
	MAE	0.35	0.45	0.39
	$R^2$	0.88	0.85	0.85
Eirunepé	RMSE	0.50	0.85	0.61
	MAE	0.39	0.70	0.50
	$R^2$	0.62	0.45	0.45
Lages	RMSE	0.61	0.77	0.62
	MAE	0.44	0.61	0.47
	$R^2$	0.83	0.83	0.83
Macapá	RMSE	0.50	0.89	0.63
	MAE	0.39	0.76	0.50
	$R^2$	0.82	0.71	0.71
Palmas	RMSE	0.60	0.72	0.67
	MAE	0.44	0.54	0.51
	$R^2$	0.87	0.77	0.77
Santa Maria	RMSE	0.59	0.67	0.65
	MAE	0.42	0.49	0.45
	$R^2$	0.88	0.86	0.86
Mean	RMSE	0.55	0.83	0.63
	MAE	0.42	0.68	0.49
	$R^2$	0.82	0.76	0.76

MARS4 - multivariate adaptive regression splines with measured data of temperature and wind speed; PMW - Penman-Monteith with measured data of temperature and wind speed. The expression "cal" indicates the locally calibrated version of a given model.

The MARS models developed with data of temperature and wind speed (MARS4), as in previous cases, had a superior performance compared to the PMW equation in its original and calibrated versions, as it obtained smaller RMSE and MAE values and higher  $R^2$  values. The highest performance improvement occurred at the Eirunepé station, where RMSE and MAE values decreased from 0.61 and 0.50 to 0.50 and 0.39, respectively, and the  $R^2$  value increased from 0.45 to 0.62 (the calibrated PMW equation was the reference equation).

### Overall evaluation of the models

It was verified that all the evaluated methods, especially those based on temperature, showed performance variations between the weather stations. This behavior can be justified by the empirical basis

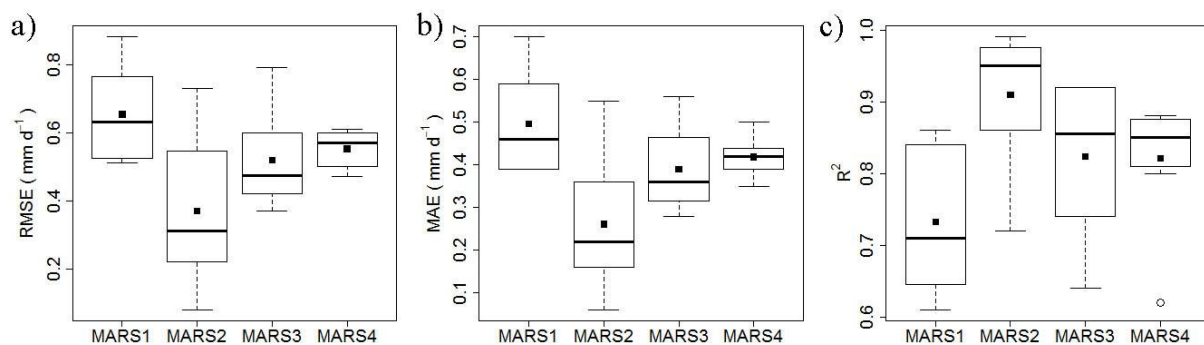


of these methods which, in turn, causes performance variations according to the climatic conditions of the location where the methods are applied (Raziei & Pereira, 2013; Feng et al., 2017).

It is also important to note the role of the calibration. In the present study, calibration improved the performance of the equations in all evaluated scenarios. Because these equations are empirical, calibration becomes a very important factor because it allows to adjust the model to the conditions of the location where it will be used. Several authors have suggested the calibration of empirical equations to obtain better estimations of  $ET_0$  (Allen et al., 1998; Liu et al., 2017; Shiri, 2017).

In all of the evaluated scenarios, MARS models were able to estimate  $ET_0$  with the best performance, surpassing conventional equations even after calibration. This behavior reaffirms the power of MARS to model complex problems, such as  $ET_0$ . Thus, the use of MARS has proven to be a viable alternative for estimation of  $ET_0$  when available climatic data are limited. Mehdizadeh et al. (2017) also reported the superiority of MARS over conventional equations for estimation of  $ET_0$ .

To compare MARS models developed in different data availability scenarios, the values of the statistical indices obtained in each scenario are presented in box plots (Figure 2).



**Figure 2.** Box plots of RMSE (a), MAE (b), and  $R^2$  (c) for the MARS models.

Based on Figure 2, it was observed that MARS2 had the best performance, followed by MARS3, MARS4, and, finally, MARS1. This behavior indicates the importance of solar radiation for estimation of  $ET_0$ , since the model that incorporated this variable (MARS2) had the best performance. The addition of relative humidity and wind speed, present in the MARS3 and MARS4 models, respectively, also promoted performance improvements in relation to the model that only used measured temperature data (MARS1), as the  $R^2$  mean value increased from 0.73 to 0.82 in both cases and RMSE and MAE values decreased. The fact that the MARS3 model had a performance slightly higher than the MARS4, which can be noted by lower RMSE and MAE mean values, indicates that, in general, relative humidity has a greater influence on  $ET_0$  at the evaluated stations than wind speed.

Corroborating with the results obtained in this study, Córdova, Carrillo-Rojas, Crespo, Wilcox, and Céleri (2015) indicated solar radiation as the most important variable for estimation of  $ET_0$ , followed by relative humidity and wind speed, in that order. Sentelhas et al. (2010) also reported a superior performance of methods using solar radiation as an input variable. In addition, according to Allen et al. (1998), solar radiation represents the largest energy source that promotes water vaporization and, consequently, evapotranspiration.

## Conclusion

Local calibration improved the performance of the evaluated equations.

The Penman-Monteith equation with missing data represents an alternative to empirical equations, having, in general, equal or superior performance.

MARS models are good options for estimation of  $ET_0$  under conditions of limited weather data, as they present the best performances across all the assessed data scenarios.

The models that used, besides temperature, solar radiation had the best performance, followed by models that used relative humidity and, finally, wind speed. The temperature-based models showed the worst performance; however, these can be used with reasonable performance, mainly by means of MARS models.

## Acknowledgements

To the National Council for Scientific and Technological Development (CNPq) for granting a scholarship to the first author and to the State Research Support Foundation of Minas Gerais State (FAPEMIG) for their financial support.

## References

- Alencar, L. P., Sedyama, G. C., & Mantovani, E. C. (2015). Estimativa da evapotranspiração de referência (ETo padrão FAO), para Minas Gerais, na ausência de alguns dados climáticos. *Engenharia Agrícola*, 35(1), 39–50. 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*. Rome, IT: FAO Irrigation and drainage paper 56.
- Almorox, J., Quej, V. H., & Martí, P. (2015). Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology*, 528(C), 514–522. DOI: 10.1016/j.jhydrol.2015.06.057
- 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. *Theoretical and Applied Climatology*, 131(1-2), 693–703. DOI: 10.1007/s00704-016-1996-2
- Antonopoulos, V. Z., & Antonopoulos, A. V. (2017). Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Computers and Electronics in Agriculture*, 132, 86–96. DOI: 10.1016/j.compag.2016.11.011
- Cheng, M.-Y., & Cao, M. (2014). Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Applied Soft Computing*, 22, 178–188. DOI: 10.1016/j.asoc.2014.05.015
- Córdova, M., Carrillo-Rojas, G., Crespo, P., Wilcox, B., & Célleri, R. (2015). Evaluation of the Penman–Monteith (FAO 56 PM) Method for Calculating Reference Evapotranspiration Using Limited Data. *Mountain Research and Development*, 35(3), 230–239. DOI: 10.1659/MRD-JOURNAL-D-14-0024.1
- Cunha, F. F., Magalhães, F. F., & Castro, M. A. (2013). Métodos para estimativa da evapotranspiração de referência para Chapadão do Sul - MS. *Engenharia na Agricultura*, 21(2), 159–172. DOI: 10.1111/j.1365
- Djaman, K., Irmak, S., & Futakuchi, K. (2017). Daily Reference Evapotranspiration Estimation under Limited Data in Eastern Africa. *Journal of Irrigation and Drainage Engineering*, 143(4), 6016015. DOI: 10.1061/(ASCE)IR.1943-4774.0001154
- Feng, Y., Cui, N., Zhao, L., Hu, X., & Gong, D. (2016). Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China. *Journal of Hydrology*, 536, 376–383. DOI: 10.1016/j.jhydrol.2016.02.053
- Feng, Y., Jia, Y., Cui, N., Zhao, L., Li, C., & Gong, D. (2017). Calibration of Hargreaves model for reference evapotranspiration estimation in Sichuan basin of southwest China. *Agricultural Water Management*, 181, 1–9. DOI: 10.1016/j.agwat.2016.11.010
- Fernandes, L. C., Paiva, C. M., & Rotunno Filho, O. C. (2012). Evaluation of six empirical evapotranspiration equations - case study: Campos dos Goytacazes/RJ. *Revista Brasileira de Meteorologia*, 27(3), 272–280. DOI: 10.1590/S0102-77862012000300002
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. DOI: 10.1214/aos/1176347963
- Gao, X., Peng, S., Xu, J., Yang, S., & Wang, W. (2015). Proper methods and its calibration for estimating reference evapotranspiration using limited climatic data in Southwestern China. *Archives of Agronomy and Soil Science*, 61(3), 415–426. DOI: 10.1080/03650340.2014.933810
- Gocic, M., Petković, D., Shamshirband, S., & Kamsin, A. (2016). Comparative analysis of reference evapotranspiration equations modelling by extreme learning machine. *Computers and Electronics in Agriculture*, 127(C), 56–63. DOI: 10.1016/j.compag.2016.05.017
- Hargreaves, G. H., & Samani, Z. A. (1985). Reference Crop Evapotranspiration from Temperature. *Applied Engineering in Agriculture*, 1(2), 96–99. DOI: 10.13031/2013.26773

- Jensen, M. E., & Haise, H. R. (1963). Estimating evapotranspiration from solar radiation. *Journal of the Irrigation and Drainage Division ASCE*, 89(1), 15-41.
- Kisi, O. (2015). Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, 528, 312-320. DOI: 10.1016/j.jhydrol.2015.06.052
- Koc, E. K., & Bozdogan, H. (2015). Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function. *Machine Learning*, 101(1-3), 35-58. DOI: 10.1007/s10994-014-5440-5
- Liu, X., Xu, C., Zhong, X., Li, Y., Yuan, X., & Cao, J. (2017). Comparison of 16 models for reference crop evapotranspiration against weighing lysimeter measurement. *Agricultural Water Management*, 184(C), 145-155. DOI: 10.1016/j.agwat.2017.01.017
- Makkink, G. F. (1957). Testing the Penman formula by means of lysimeters. *Journal Inst. Water Engineering*, 11(3), 277-288.
- Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2017). Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and Electronics in Agriculture*, 139, 103-114. DOI: 10.1016/j.compag.2017.05.002
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, 303(1-4), 290-306. DOI: 10.1016/j.jhydrol.2004.08.026
- Pereira, L. S., Allen, R. G., Smith, M., & Raes, D. (2015). Crop evapotranspiration estimation with FAO56: Past and future. *Agricultural Water Management*, 147, 4-20. DOI: 10.1016/j.agwat.2014.07.031
- Raziei, T., & Pereira, L. S. (2013). Estimation of ETo with Hargreaves-Samani and FAO-PM temperature methods for a wide range of climates in Iran. *Agricultural Water Management*, 121, 1-18. DOI: 10.1016/j.agwat.2012.12.019
- Sentelhas, P. C., Gillespie, T. J., & Santos, E. A. (2010). Evaluation of FAO Penman-Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. *Agricultural Water Management*, 97(5), 635-644. DOI: 10.1016/j.agwat.2009.12.001
- Shiri, J. (2017). Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran. *Agricultural Water Management*, 188(C), 101-114. DOI: 10.1016/j.agwat.2017.04.009
- Tabari, H., Grismer, M. E., & Trajkovic, S. (2013). Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science*, 31(2), 107-117. DOI: 10.1007/s00271-011-0295-z
- Tabari, H., Kisi, O., Ezani, A., & Talaei, P. H. (2012). SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology*, 444-445, 78-89. DOI: 10.1016/j.jhydrol.2012.04.007
- Talaei, P. H. (2014). Performance evaluation of modified versions of Hargreaves equation across a wide range of Iranian climates. *Meteorology and Atmospheric Physics*, 126(1-2), 65-70. DOI: 10.1007/s00703-014-0333-5
- Valiantzas, J. D. (2013). Simplified forms for the standardized FAO-56 Penman-Monteith reference evapotranspiration using limited weather data. *Journal of Hydrology*, 505, 13-23. DOI: 10.1016/j.jhydrol.2013.09.005
- Yassin, M. A., Alazba, A. A., & Mattar, M. A. (2016). Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agricultural Water Management*, 163(C), 110-124. DOI: 10.1016/j.agwat.2015.09.009