Computers and Electronics in Agriculture 165 (2019) 104937

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Empirical and learning machine approaches to estimating reference evapotranspiration based on temperature data

Matheus Mendes Reis[a,*], Ariovaldo José da Silva[a], Jurandir Zullo Junior[a], Leonardo David Tuffi Santos[b], Alcinei Místico Azevedo[b], Érika Manuela Gonçalves Lopes[b]

[a] School of Agricultural Engineering, University of Campinas, Avenida Cândido Rondon, 501, Barão Geraldo, 13083-875 Campinas, SP, Brazil
[b] Institute of Agrarian Sciences, Federal University of Minas Gerais, Avenida Universitária, 1000, Bairro Universitário, 39404-547 Montes Claros, MG, Brazil

ARTICLE INFO

Keywords:
Soft computing
Artificial neural networks
Multiple linear regression
Extreme learning machines
Artificial intelligence
Meteorological data

ABSTRACT

The precise estimation of reference evapotranspiration (ET_0) is crucial for the planning and management of water resources and agricultural production. In this study, the applicability of the Hargreaves Samani (HS), artificial neural network (ANN), multiple linear regression (MLR) and extreme learning machine (ELM) models were evaluated to estimate ET_0 based on temperature data from the Verde Grande River basin, southeastern Brazil. These models were evaluated in two scenarios: local and pooled. In the local scenario, training, calibration and validation of the models were performed separately at each station. In the pooled scenario, meteorological data from all stations were grouped for training and calibration and then separately tested at each station. The ET_0 values estimated by the Penman-Monteith model (FAO-56 PM) were considered the target data. All the developed models were evaluated by cluster analysis and the following performance indices: relative root mean square error (RRMSE), Pearson correlation coefficient (r) and Nash-Sutcliffe coefficient (NS). In both scenarios evaluated, local and pooled, the results revealed the superiority of the artificial intelligence methods (ANN and ELM) and the MLR model compared to the original and adjusted HS models. In the local scenario, the ANN (with r of 0.751, NS of 0.687 and RRMSE of 0.112), ELM (with r of 0.747, NS of 0.672 and RRMSE of 0.116) and MLR (with r of 0.743, NS of 0.665 and RRMSE of 0.068) models presented the best performance, in addition to being grouped in the same cluster. Similar to the observations from the local scenario, the ANN (with r of 0.718, NS of 0.555 and RRMSE of 0.165), ELM (with r of 0.724, NS of 0.601 and RRMSE of 0.151) and MLR (with r of 0.731, NS of 0.550 and RRMSE of 0.091) models presented the best performance in the pooled scenario and were grouped in the same cluster. The locally trained models presented higher precision than the models generated with pooled data; however, the models generated in the pooled scenario could be used to estimate ET_0 in cases of unavailability of local meteorological data. Although the MLR, ANN and ELM models, based on temperature data, are appropriate alternatives to accurately estimate ET_0 in the Verde Grande River basin, southeastern Brazil, the MLR model presents the advantage of the use of explicit algebraic equations, facilitating its application.

1. Introduction

The reference evapotranspiration (ET_0), introduced by the Food and Agriculture Organization of the United Nations (FAO) as a methodology for computing crop evapotranspiration (Doorenbos and Pruitt, 1977), is an essential component in irrigation planning, river basin hydrology, and hydrological balance studies (Antonopoulos and Antonopoulos, 2017; Traore et al., 2010). In addition, ET_0 is a key element in executing effective water management practices and optimizing their use in agricultural production areas (Smith, 2000). In the case of arid and semi-arid regions, understanding ET_0 is even more important for efficient irrigation planning (Huo et al., 2012).

ET_0 can be determined by lysimeters (Anapalli et al., 2016; Xu et al., 2018), the energy balance (Yan et al., 2017), scintillometers (Valayamkunnath et al., 2018), or by using empirical equations based on meteorological data (Antonopoulos and Antonopoulos, 2017). The Penman-Monteith (PM) method is universally recommended by the FAO as the only precise equation to calculate ET_0 (Allen et al., 1998). The PM model incorporates thermodynamic and aerodynamic aspects, and it has been shown to be relatively accurate in humid and arid

* Corresponding author.
E-mail addresses: matheussmendes@hotmail.com, m191053@dac.unicamp.br (M.M. Reis).

https://doi.org/10.1016/j.compag.2019.104937
Received 15 April 2019; Received in revised form 31 July 2019; Accepted 1 August 2019
Available online 14 August 2019
0168-1699/

Table 1
Meteorological stations and statistical properties of the meteorological variables of the study area.

Station (longitude; latitude and altitude)	Station code	Variable	Min	Max	Mean	SD	CV (%)
Espinosa (42.8°W, 14.9°S and 569.6 m)	1	Pr (mm)	0.0	126.7	1.9	7.7	411.2
		RH (%)	22.5	97.7	63.4	13.7	21.6
		Tmin (°C)	9.6	27.7	20.0	2.6	13.0
		Tmax (°C)	20.2	40.3	31.3	2.8	8.9
		U_2(m s^{-1})	0.0	6.4	2.2	1.1	49.3
		Rs (MJ m^{-2})	7.5	76.6	21.9	4.9	22.3
		ET_0 (mm d^{-1})	1.5	11.5	4.9	1.3	26.3
Janaúba (43.3°W, 15.8°S and 516.0 m)	2	Pr (mm)	0.0	157.0	2.0	8.1	402.5
		RH (%)	21.0	97.8	59.1	13.9	23.6
		Tmin (°C)	7.8	29.0	19.3	2.5	13.0
		Tmax (°C)	20.2	41.8	32.1	2.9	9.1
		U_2(m s^{-1})	0.0	7.5	1.1	1.0	94.9
		Rs (MJ m^{-2})	7.3	36.5	21.4	4.5	20.8
		ET_0 (mm d^{-1})	1.6	10.0	4.3	1.1	24.6
Juramento (43.7°W, 16.8°S and 650.0 m)	3	Pr (mm)	0.0	127.0	2.4	8.5	348.2
		RH (%)	24.0	99.8	69.6	12.5	17.9
		Tmin (°C)	4.2	25.3	16.8	3.2	19.3
		Tmax (°C)	19.7	40.0	29.9	2.8	9.5
		U_2(m s^{-1})	0.0	6.5	1.1	0.5	45.1
		Rs (MJ m^{-2})	7.0	29.9	20.6	4.6	22.5
		ET_0 (mm d^{-1})	1.3	7.8	3.9	1.0	25.0
Montes Claros (43.8 °W, 16.7 °S and 646.3 m)	4	Pr (mm)	0.0	145.0	2.7	9.3	344.0
		RH (%)	24.0	98.3	62.7	14.6	23.4
		Tmin (°C)	5.8	28.0	18.0	3.0	16.4
		Tmax (°C)	20.0	40.3	30.2	2.9	9.5
		U_2(m s^{-1})	0.0	4.0	1.4	0.5	33.5
		Rs (MJ m^{-2})	7.1	30.1	20.9	5.0	24.0
		ET_0 (mm d^{-1})	1.4	8.5	4.2	1.1	26.3
Monte Azul (42.8 °W; 15.1 °S and 603.6 m)	5	Pr (mm)	0.0	140.1	2.1	7.8	381.3
		RH (%)	21.0	98.8	58.6	14.3	24.4
		Tmin (°C)	11.3	28.0	20.4	2.2	11.0
		Tmax (°C)	20.7	41.1	31.3	2.9	9.2
		U_2(m s^{-1})	0.0	6.7	1.7	1.1	60.3
		Rs (MJ m^{-2})	6.4	29.7	21.4	4.6	21.5
		ET_0 (mm d^{-1})	1.4	9.4	4.8	1.2	24.9

Pr – precipitation; RH – relative humidity; Tmin – minimum temperature; Tmax – maximum temperature; U_2 – wind speed; Rs – solar radiation; ET_0 – reference evapotranspiration by Penman-Monteith (Allen et al., 1998); Min – minimum; Max – maximum; SD – standard deviation; CV (%) – coefficient of variation.

regions (Smith et al., 1991; Yin et al., 2008). However, the greatest disadvantage of the PM method is the need for several types of climatic data that are not always available (Cobaner et al., 2017; Fan et al., 2018; Feng et al., 2017; Gocic et al., 2015).

In the Brazilian semi-arid region, the distribution and density of the meteorological stations are inadequate or insufficient, limiting the use of the PM method in irrigation management in this region, which has the largest public irrigated perimeter in Latin America and other important irrigation perimeters.

The Hargreaves Samani (HS) method is an alternative to the PM equation when the data set required by the PM model is not fully available (Allen et al., 1998; Cobaner et al., 2017). As shown by Almorox et al. (2015), the HS equation shows a more accurate performance in arid, semi-arid, temperate, cold and polar regions compared to other empirical models that rely on air temperature data to estimate ET_0. However, this model overestimates ET_0 under high relative humidity conditions and underestimates it under conditions where the wind speed is higher than 3 m s^{-1} (Allen et al., 1998; Didari and Ahmadi, 2019; Droogers and Allen, 2002). In this situation, the local calibration of empirical models or the creation of local equations is necessary.

In recent years, artificial intelligence (AI) methods, or soft computing methods, have been successfully applied to estimate and predict ET_0 in limited weather data situations. Artificial neural networks (ANNs) (Antonopoulos and Antonopoulos, 2017; Kumar et al., 2002; Traore et al., 2016; Traore et al., 2010), adaptive neuro-fuzzy inference systems (ANFIS) (Citakoglu et al., 2014; Dou and Yang, 2018), genetic programming (GP) (Kisi et al., 2015; Kisi and Sanikhani, 2015), support vector machines (SVM), wavelet neural networks (Cobaner, 2013; Gocic et al., 2015; Fan et al., 2018; Kisi and Cimen, 2009), tree-based assemble algorithms (i.e., random forest (RF), M5 model tree (M5Tree), gradient boosting decision tree (GBDT) and extreme gradient boosting (XGBoost)) (Fan et al., 2018; Feng et al., 2017a), generalized regression neural networks (GRNN) (Feng et al., 2017a) and extreme learning machines (ELM) (Dou and Yang, 2018; Feng et al., 2017b; Gocic et al., 2016) are examples of proposed AI models for the successful estimation of ET_0 in regions with low availability of meteorological data.

The Verde Grande River basin is one of the main irrigated fruit growing regions in Brazil and stands out because it has the largest irrigated perimeter in Latin America; however, the low availability of meteorological stations in the region hampers the accurate estimation of ET_0. The development of improved methods to estimate the amount of water required by the crops is essential to improve the accuracy of the irrigation level and to increase the efficiency of water use (Antonopoulos and Antonopoulos, 2017), especially in areas with water scarcity problems, such as the semi-arid region, where the Verde Grande River basin is located. The low spatial density of meteorological stations in the Verde Grande River basin and its poor distribution make it impossible to apply the PM equation in irrigation management in most of the properties. Estimating calibrated models for the determination of ET_0 in the Verde Grande River basin using only temperature data can result in advances in irrigation management and regional hydrological studies, as well as a more economical alternative to the installation of new, spatially close meteorological stations.

The main objective of this study was to evaluate the applicability of different methods in estimating daily reference evapotranspiration

(ET_0) only with air temperature data as the input, thus allowing the accurate determination of ET_0 in locations without available relative air humidity, solar radiation and wind speed data. To achieve this purpose, Hargreaves Samani (HS), adjusted HS, artificial neural network "multilayer perceptron" (ANN), multiple linear regression (MLR) and extreme learning machine (ELM) were compared to the Penman-Monteith method (FAO 56 – PM). These models were evaluated in two scenarios: local and pooled. In the local scenario, all models were trained, calibrated and validated separately at each station. In the pooled scenario, the meteorological data from all stations were grouped for training and calibration of the models and then tested separately at each station. The generation of models by a combined approach allows the estimation of ET_0 in cases of unavailability of local meteorological data, that is, in regions that are distant from meteorological stations, but with the availability of temperature data.

2. Material and methods

2.1. Study area and data set

The Verde Grande River basin has an area of approximately 31,410 km^2 and a population of 741,500 inhabitants. The Verde Grande River basin stands out in the world scenario for housing the largest irrigated perimeter in Latin America, with the irrigation projects Jaíba, Gorutuba, Lagoa Grande and Estreito, and an irrigated area of approximately 742 km^2. The study area is located in the semi-arid region of Brazil. The climate of the region is classified, according to Köppen, as Aw, warm tropical with a dry winter. The meteorological stations and statistical properties of the climatic variables are shown in Table 1.

The daily meteorological variables (maximum (Tmax) and minimum (Tmin) air temperature at a height of 2 m, mean relative humidity (RH), wind speed at a height of 10 m (U_{10}) and sunshine duration) were obtained from the five meteorological stations of the Instituto Nacional de Meteorologia (INMET) located within the Verde Grande River basin (Fig. 1) between 1996 and 2016. The meteorological data provided by INMET are of satisfactory quality certified by ISO 9001: 2008. (INMET, 2019). The missing data for Tmax and Tmin were reconstituted by the linear interpolation method, and the missing data for RH, U_2 and Rs were estimated according to the methodology proposed by Allen et al. (1998).

2.2. FAO-56. Penman-Monteith model

The FAO-recommended Penman-Monteith equation (FAO 56-PM) (Equation (1)) (Allen et al., 1998) was used to estimate ET_0 data, which were used as the targets for the calibration and evaluation of the HS, adjusted HS, MLR, ANN and, ELM models. This process is an accepted and commonly used practice (Antonopoulos and Antonopoulos, 2017; Didari and Ahmadi, 2019; Dou and Yang, 2018).

$$ET_0 = \frac{0.408\Delta(Rn - G) + \gamma\frac{900}{T + 273}u_2(es - ea)}{\Delta + \gamma(1 + 0.34u_2)} \tag{1}$$

where ET $_0$ is the reference evapotranspiration (mm dia^{-1}), Rn is the net radiation (MJ m^{-2} day^{-1}), G is the soil heat flux density (MJ m^{-2} day^{-1}), T is the mean daily air temperature at a height of 2 m (°C), U_2 is the wind speed at a height of 2 m (m s^{-1}), es is the saturated vapor pressure (kPa), ea is the actual vapor pressure (kPa), Δ is the slope vapor pressure curve (kPa $°C^{-1}$), and γ is the psychrometric constant (kPa $°C^{-1}$).

Due to the lack of U_2 and Rs data, these two parameters were estimated (Eqs. (2) and (3)) based on the data for sunshine duration and U_{10}, respectively (Allen et al., 1998).

$$U_2 = U_{10}\frac{4.87}{\ln 67.8z - 5.42} \tag{2}$$

$$R_S = \left(a_S + b_S\frac{n}{N}\right)R_a \tag{3}$$

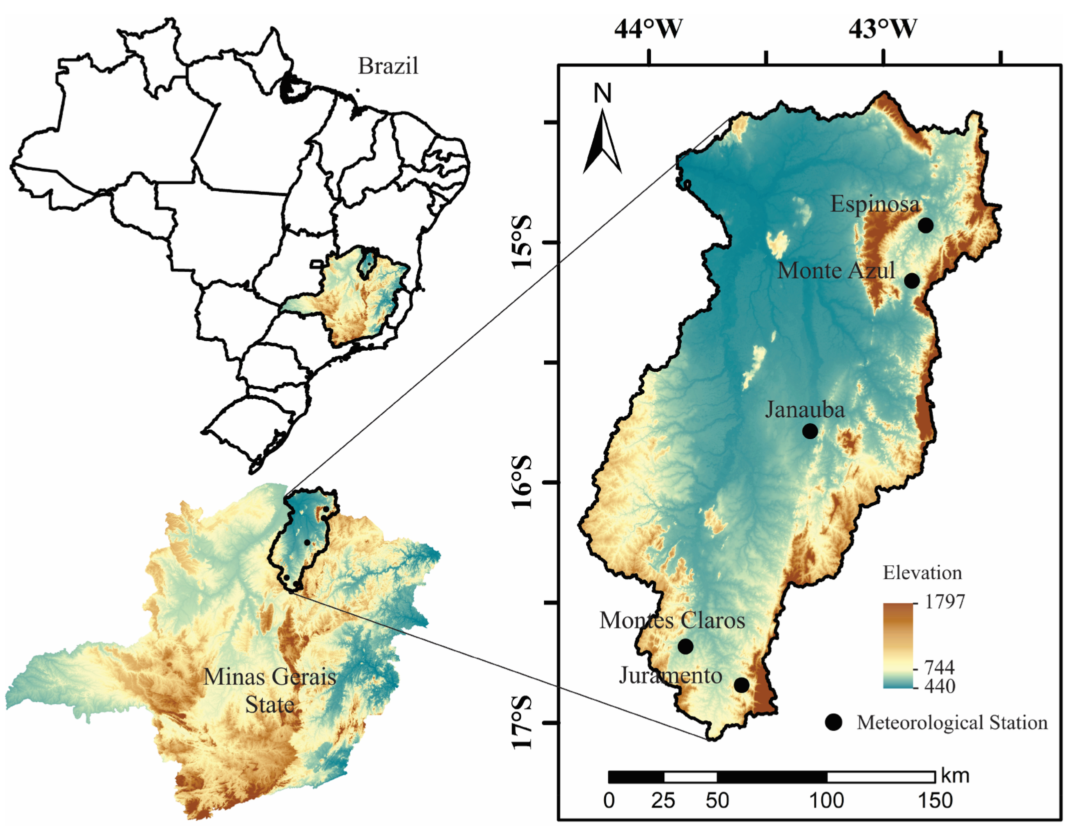

Fig. 1. Geographical location of the meteorological stations.

Table 2
Multiple regression models tested to estimate ET_0 (Zi) as a function of maximum (xi) and minimum (yi) temperatures.

Models	Function
1	$Z_i = a + bx_i + cy_i + e_i$
2	$Z_i = a + bx_i + cx_i2 + dy_i + e_i$
3	$Z_i = a + bx_i + cy_i + dy_i^2 + e_i$
4	$Z_i = a + bx_i + cx_i^2 + dy_i + fy_i^2 + e_i$
5	$Z_i = a + bx_i + cy_i + dx_iy_i + e_i$
6	$Z_i = a + bx_i + cx_i2 + dy_i + fx_iy_i + e_i$
7	$Z_i = a + bx_i + cy_i + dy_i^2 + fx_iy_i + e_i$
8	$Z_i = a + bx_i + cx_i^2 + dy_i + fy_i^2 + gx_iy_i + e_i$
9	$Z_i = a + bx_i + cx_i^2 + dy_i + fy_i^2 + gx_iy_i + hx_i^2y_i + e_i$
10	$Z_i = a + bx_i + cx_i^2 + dy_i + fy_i^2 + gx_iy_i + hx_iy_i^2 + e_i$
11	$Z_i = a + bx_i + cx_i^2 + dy_i + fy_i^2 + gx_iy_i + hx_i^2y_i + jx_iy_i^2 + e_i$
12	$Z_i = a + bx_i + cx_i^2 + dy_i + fy_i^2 + gx_iy_i + hx_i^2y_i + jx_iy_i^2 + k_i^2y_i^2 + e_i$

where U_{10} is the wind speed at a height of 10 m ($m\,s^{-1}$), z is the height measurement (10 m), Rs is the solar or shortwave radiation ($MJ\,m^{-2}\,day^{-1}$), n is the sunshine duration (h), N is the maximum possible sunshine or daylight duration (h), R_a is extraterrestrial radiation ($MJ\,m^{-2}\,dia^{-1}$), and a_s and b_s are constants with a value of 0.28 and 0.52, respectively, as recommended by FAO-56 (Allen et al., 1998).

2.3. Hargreaves Samani model

The Hargreaves Samani (HS) model (Eq. (4)) was initially proposed by Hargreaves and Samani (1985) and requires only air temperature data to estimate ET_0.

$$ET_0 = 0.0023R_a\,(\text{Tmax} - \text{Tmin})^{0.5}(\text{T} + 17.8) \tag{4}$$

where T max and Tmin are the maximum and minimum air temperatures (°C), respectively.

The extraterrestrial radiation data (Ra) were calculated based on latitude data (Equation (5)).

$$R_a = \frac{24(60)}{\pi}G_{SC}\,d_r\,[\omega_S\sin(\varphi)\sin(\delta) + \cos(\varphi)\cos(\delta)\cos(\omega_S)] \tag{5}$$

where R_a is the extraterrestrial radiation ($MJ\,m^{-2}d^{-1}$), G_{SC} is the solar constant (0.0820 $MJ\,m^{-2}\,min^{-1}$), dr is the inverse relative Earth-Sun distance, ω_S is the sunset hour angle (rad), φ is the latitude (rad) and δ is the solar declination (rad).

The adjusted HS model was obtained by regression (Eq. (6)), which is an accepted and commonly used practice (Droogers and Allen, 2002; Feng et al., 2017b; Shiri et al., 2014).

$$ET_0^{PM} = a + bET_0^{HS} \tag{6}$$

where ET_0^{PM} is ET_0 estimated by FAO-56 PM, ET_0^{HS} is ET_0 estimated by

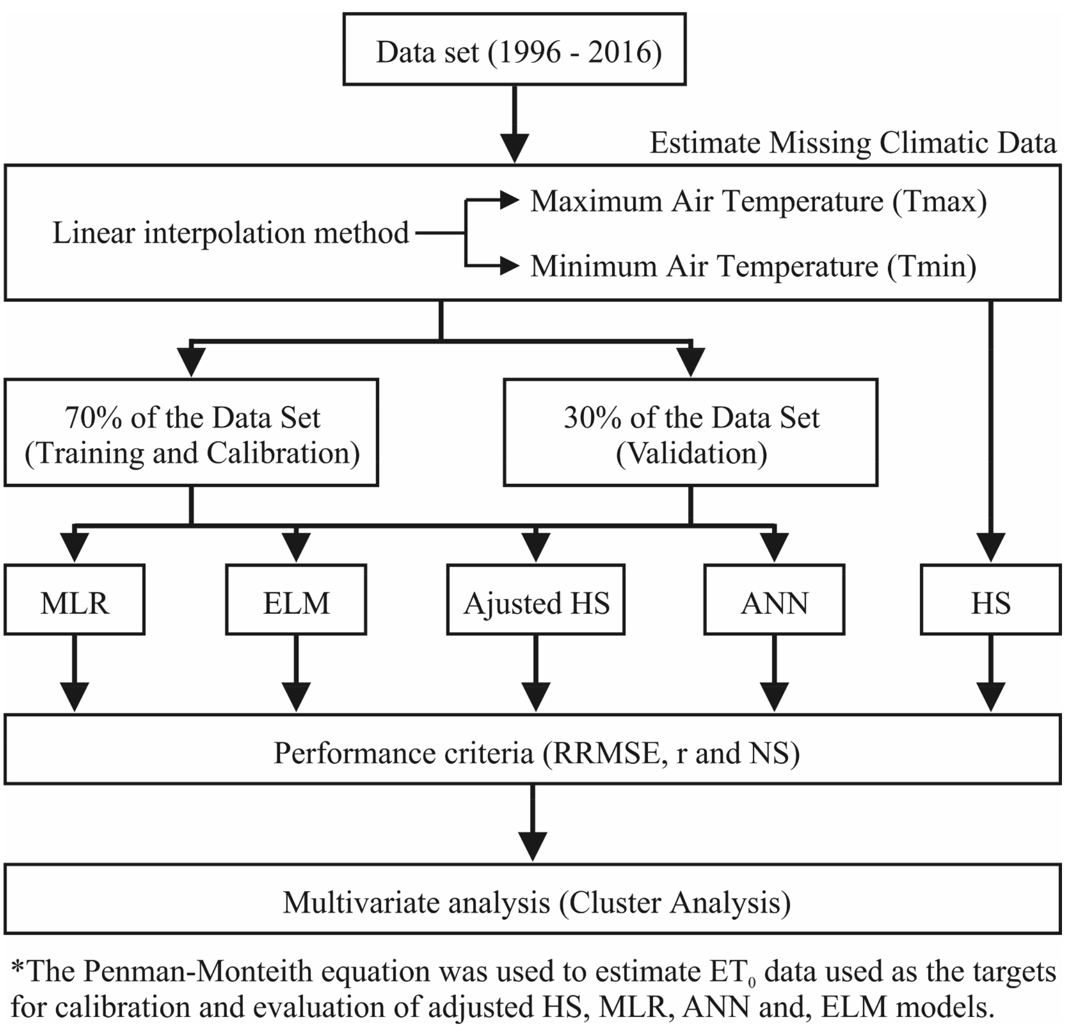

Fig. 2. Workflow of the ET_0 estimation in this study. MLR – multiple linear regression; ELM – extreme learning machine; HS – Hargreaves Samani; ANN – artificial neural Network "multilayer perceptron"; RRMSE - relative root mean square error; r - Pearson correlation coefficient; NS – Nash-Sutcliffe coefficient.

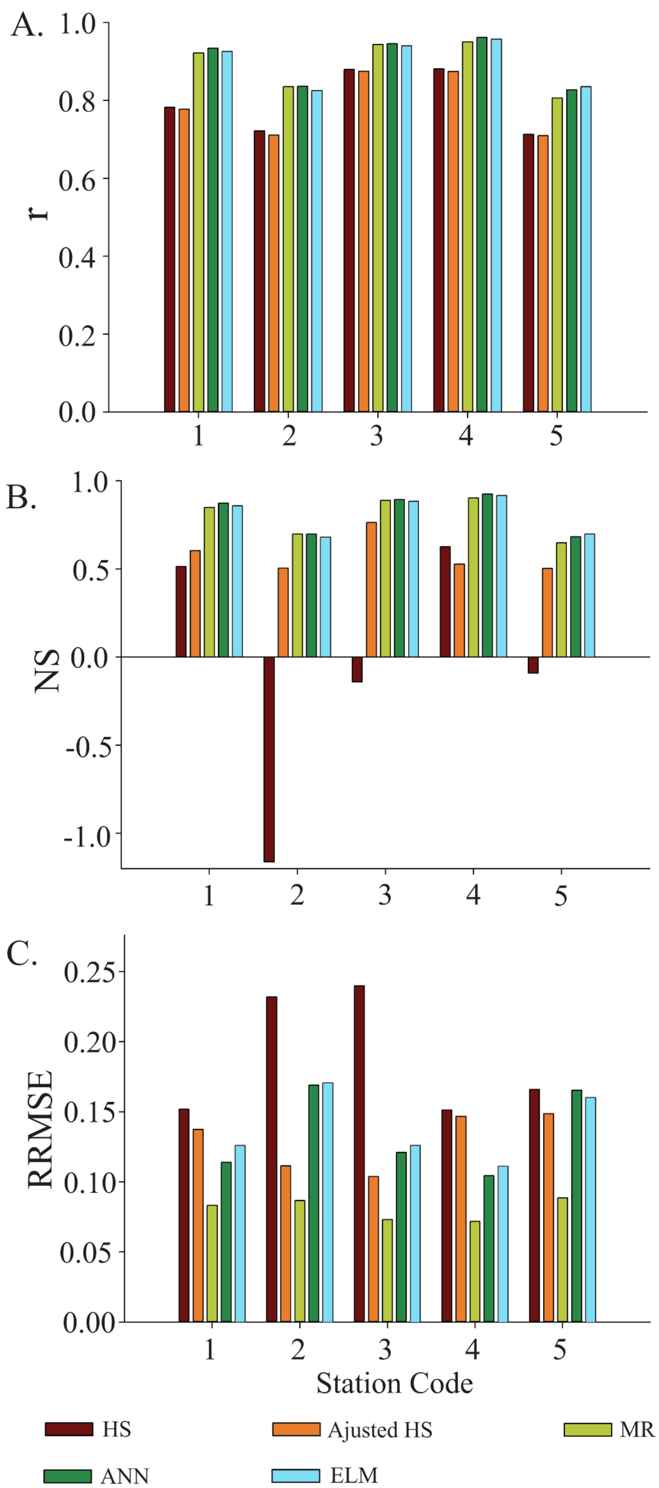

Fig. 3. Pearson correlation coefficient (r) (A), Nash-Sutcliffe coefficient (NS) (B) and relative root mean square error (RRMSE) (C) of the models during the study period (1996 to 2016) in the local scenario.

HS, and a and b are regression coefficients.

2.4. Artificial neural networks

Artificial Neural Networks (ANNs) are mathematical models that are analogous to biological neural networks, which have been applied in many studies to model ET_0 (Dou and Yang, 2018; Traore et al., 2010).

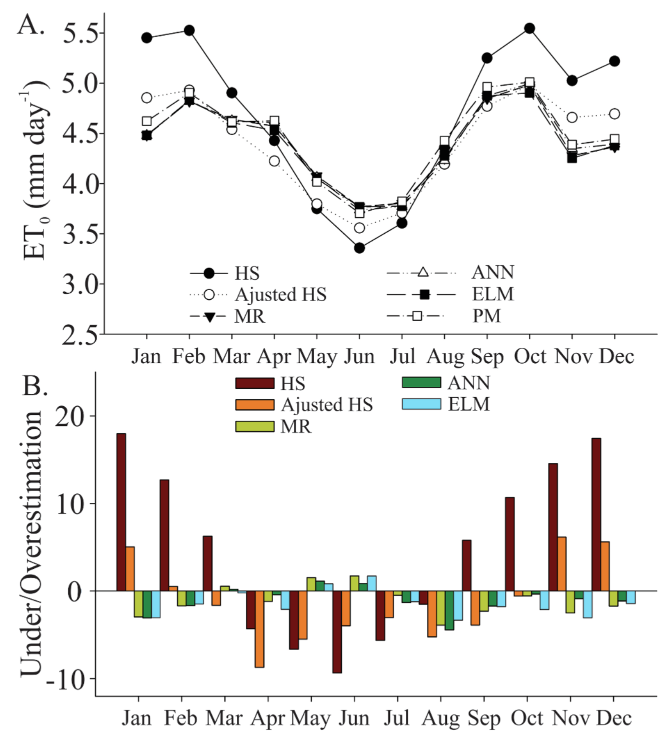

Fig. 4. Monthly variation in ET_0 values estimated by PM, HS, adjusted HS, MLR, ANN and ELM models generated with local data (A) and under/overestimation values of HS, adjusted HS, MLR, ANN and ELM models in relation to the PM model (B).

Multilayer "perceptron" (MLP) networks have been used with "backpropagation" learning. For the development of MLP networks, the *mlp* function of the RSNNS package in the R software was used, with the *backpropagation* algorithm (*learnFunc* = *"Std_Backpropagation"*) and a learning rate of 0.1 (*learnFuncParams* = 0.1). The networks were composed of an input layer with two neurons, corresponding to the number of input variables (Tmax and Tmin). In the output layer, a neuron corresponding to ET_0 was introduced. To define the number of neurons in the hidden layer, a trial and error procedure was used. For this procedure, 1000 networks were tested with a number of neurons in the middle layer ranging from 1 to 10. The most appropriate model for the studied five cities was the one that presented eight neurons in the hidden layer, presenting smaller estimates of the mean square error.

The common sigmoid and linear activation functions were used for the hidden and output layers, respectively. The number of training times was arbitrated as 500.

2.5. Multiple linear regression

To estimate ET_0 from the maximum and minimum temperature, regression was also used. For this procedure, 12 regression models were tested (Table 2), and the quality of the fit was evaluated by estimating the Akaike information criterion (AIC). To adjust the regression models, we used the lm function in software R.

2.6. Extreme learning machine

The ELM learning machine technique was initially proposed by Huang et al. (2006), and it has been applied in several studies on ET_0 estimation (Dou and Yang, 2018; Feng et al., 2017b; Gocic et al., 2016) and in other research areas (Cao et al., 2016; Duan et al., 2016). ELM has an extremely fast learning speed in comparison to other learning machine techniques such as ANN. In addition, the number of neurons in the hidden layer for a specific ELM model does not need to be obtained

Table 3
Algebraic equations generated by multiple linear regression (MLR) and Hargreaves Samani (HS) models adjusted to estimate ET_0 in the local scenario.

City	Algebraic equation			
	MLR	R^2	Adjusted HS	R^2
Espinosa	$ET_0 = -42.3^{ns} + 2.564^{ns}(Tmax) - 0.04672^{ns}(Tmax^2) + 5.242^{*}(Tmin) - 0.2121^{**}(Tmin^2) - 0.259^{ns}(Tmax)(Tmin) + 0.004269^{ns}(Tmax^2)(Tmin) + 0.01022^{*}(Tmax)(Tmin^2) - 0.0001443^{*}(Tmax^2)(Tmin^2)$	0.85	$ET_0 = 1.22735^{***} + 0.78699^{***}ET_0^{HS}$	0.59
Janauba	$ET_0 = 23.761^{***} - 1.4303^{***}(Tmax) + 0.022346^{***}(Tmax^2) - 1.3629^{***}(Tmin) - 0.005278^{***}(Tmin^2) + 0.09384^{***}(Tmax)(Tmin) - 0.0013245^{***}(Tmax^2)(Tmin)$	0.70	$ET_0 = 2.25225^{***} + 0.42465^{***}ET_0^{HS}$	0.50
Juramento	$ET_0 = 10.329^{***} - 0.43629^{**}(Tmax) + 0.004605^{ns}(Tmax^2) - 1.326^{***}(Tmin) + 0.0159^{**}(Tmin^2) + 0.06947^{***}(Tmax)(Tmin) - 0.000544^{***}(Tmax^2)(Tmin) - 0.000754^{***}(Tmax)(Tmin^2)$	0.89	$ET_0 = 0.63383^{***} + 0.69692^{***}ET_0^{HS}$	0.76
Montes Claros	$ET_0 = 4.059^{*} + 0.1218^{ns}(Tmax) - 0.0111^{***}(Tmax^2) - 1.332^{***}(Tmin) + 0.02688^{*}(Tmin^2) + 0.0694^{***}(Tmax)(Tmin) - 0.001942^{***}(Tmax)(Tmin^2) + 0.0000113^{ns}(Tmax^2)(Tmin^2)$	0.90	$ET_0 = 0.18561^{***} + 0.87316^{***}ET_0^{HS}$	0.76
Monte Azul	$ET_0 = 94.72^{**} - 5.948^{**}(Tmax) + 0.09489^{**}(Tmax^2) - 8.301^{*}(Tmin) + 0.1764^{*}(Tmin^2) + 0.5284^{*}(Tmax)(Tmin) - 0.008155^{*}(Tmax^2)(Tmin) - 0.01113^{*}(Tmax)(Tmin^2) + 0.0001712^{*}(Tmax^2)(Tmin^2)$	0.68	$ET_0 = 2.48186^{***} + 0.50762^{***}ET_0^{HS}$	0.51

Tmax – maximum temperature; Tmin – minimum temperature; R^2 – coefficient of determination; ns – not significant by Student's *t* test; and *, **, *** significant at levels of 0.05, 0.01 and 0.001 by Student's *t* test, respectively.

Table 4
Global average performance and annual average of ET_0 values estimated by the HS, adjusted HS, MLR, ANN and ELM models generated with local data.

Model	r	NS	RRMSE	ET0 (mm)	Under/Overestimation (%)
HS	0.663	−0.042	0.157	1714.7	4.8
Adjusted HS	0.658	0.484	0.101	1608.3	−1.3
MLR	0.743	0.665	0.068	1608.1	−1.1
ANN	0.751	0.678	0.112	1609.5	−1.1
ELM	0.747	0.672	0.116	1603.0	−1.4
PM	/	/	/	1627.5	/

r – Pearson correlation coefficient; NS – Nash-Sutcliffe coefficient; RRMSE – relative root mean square error; ET_0 – reference evapotranspiration; HS – Hargreaves Samani; MLR – multiple linear regression; ANN – artificial neural networks; ELM – extreme learning machine; PM – Penman-Monteith.

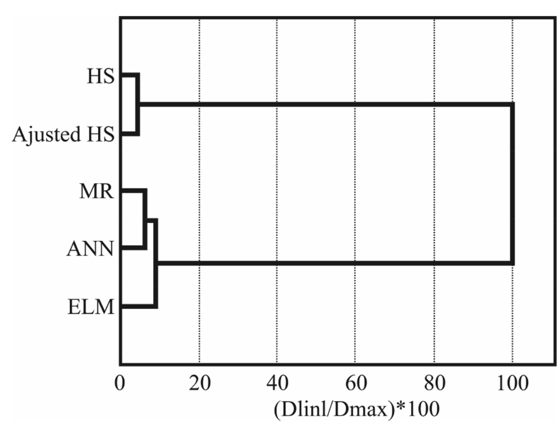

Fig. 5. Hierarchical dendrogram obtained by the Ward algorithm for the HS, adjusted HS, MLR, ANN and ELM models generated with local data.

by a trial and error procedure, as in the case of models generated by ANNs. In addition, adjustment of all the parameters (e.g., the input weights and biases) in relation to the hidden neurons is not required because these parameters are independent of the training data. More details about the ELM model can be found in Huang et al. (2006) and Gocic et al. (2016).

The ELM model consists of an input layer with two neurons (Tmax and Tmin), a single hidden layer with six neurons and the output layer (ET_0). For the generation of ELM models, the *elmtrain* function of the *elmNN* package in R software was used. The symmetric saturating linear transfer function and the linear function were used in the hidden and output layers, respectively.

2.7. Calibration and validation of the models

In this study, 70% of the data set was used for training and calibration, while the remaining 30% was used for validation of the adjusted HS, MLR, ANN and ELM models. This division was performed through the *splitForTrainingAndTest* function of the RSNNS package in R software. To ensure that the data set was randomly distributed between the training set and the validation set, the data were previously randomized using the *sample* function of the *base* package in R software.

All the input data from the ELM and ANN models were normalized (ranging from 0 to 1) to minimize the influence of the absolute scale. The normalization scheme is defined by Eq. (7).

$$x_{norm} = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \tag{7}$$

where x_{norm}, x_0, x_{min} and x_{max} are the normalized value, real value, minimum value and maximum value, respectively.

For normalization, the *normalizeData* function of the RSNNS package in R software was used.

The normalized data for ET_0 obtained by the ANN and ELM models were transformed into mm day^{-1} unit using Eq. (8).

$$V_{dn} = [V_{max} + (V_n - 1)(V_{max} - V_{min})] \tag{8}$$

where V_{dn} is the denormalized value (dimensionless), V_n is the normalized value, V_{min} is the minimum value of the sample, and V_{max} is the maximum value of the sample.

The ANN and ELM models were tested 1000 times, storing their respective mean squared errors (MSE) for the validation data set. The MSE for the validation sample was determined using Eq. (9).

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n} \tag{9}$$

where n is the total number of data tested, and P_i and O_i are the ET_0 values obtained by the FAO-56 PM method and by the artificial intelligence methods (ANN and ELM), respectively.

The best fit network for the ANN and ELM models was established by the smaller MSE for the validation sample, with the objective to ensure the absence of overfitting.

2.8. Performance criteria

The relative root mean square error (RRMSE) (Eq. (10)), Pearson correlation coefficient (r) (Eq. (11)) and Nash-Sutcliffe coefficient (NS) (Eq. (12)) were used to evaluate the performance of the HS, adjusted HS, MLR, ANN and ELM models.

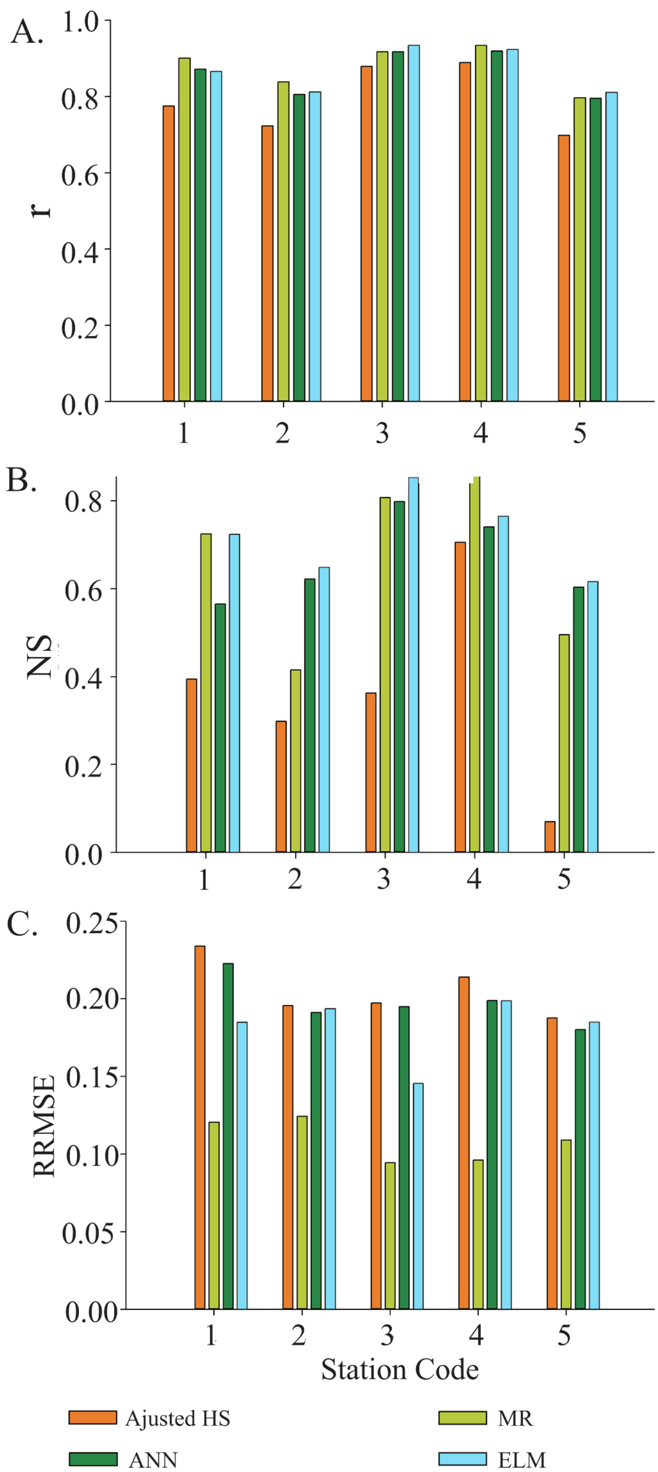

Fig. 6. Pearson correlation coefficient (r) (A), Nash-Sutcliffe (NS) coefficient (B) and relative root mean square error (RRMSE) (C) of the adjusted HS, MLR, ANN and ELM models generated with pooled data.

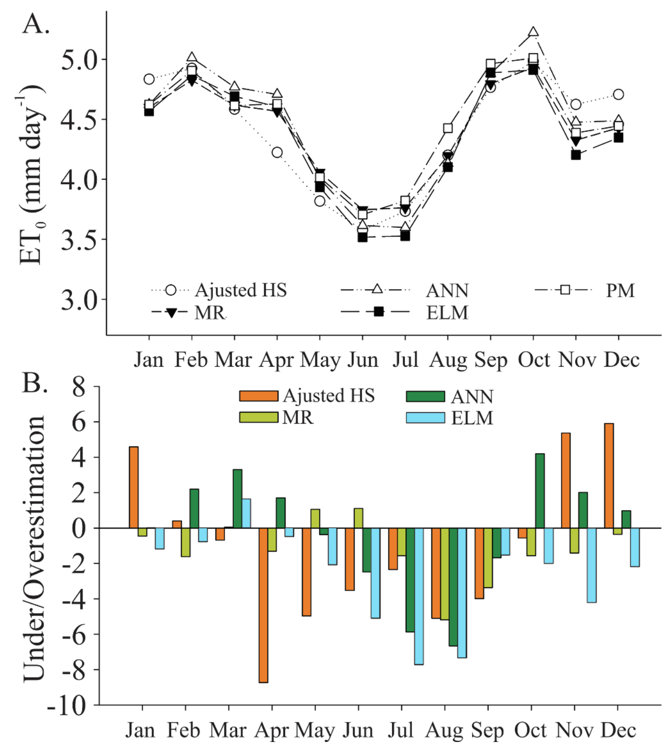

Fig. 7. Monthly variation of ET_0 values estimated by PM, HS, adjusted HS, MLR, ANN and ELM models generated with pooled data (A), and under/overestimation values of HS, adjusted HS, MLR, ANN and ELM models in relation to the PM model (B).

Table 6
Global average performance and annual average of ET_0 values estimated by adjusted HS, MLR, ANN and ELM models generated with pooled data.

Model	r	NS	RRMSE	ET0 (mm)	Under/Overestimation (%)
Adjusted HS	0.661	0.305	0.171	1610.1	−1.1
MLR	0.731	0.550	0.091	1606.7	−1.2
ANN	0.718	0.555	0.165	1626.4	−0.2
ELM	0.724	0.601	0.151	1585.1	−2.7
PM	/	/	/	1627.5	/

r – Pearson correlation coefficient; NS – Nash-Sutcliffe coefficient; RRMSE – relative root mean square error; ET_0 – reference evapotranspiration; HS – Hargreaves Samani; MLR – multiple linear regression; ANN – artificial neural networks; ELM – extreme learning machine; PM – Penman-Monteith.

$$RRMSE = \frac{RMSE}{P_i} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}}{\bar{P}_i} \quad (10)$$

$$r = \frac{n(\sum_{i=1}^{n} O_i.\, P_i) - (\sum_{i=1}^{n} O_i).\,(\sum_{i=1}^{n} P_i)}{\sqrt{(n\sum_{i=1}^{n} O_i^2 - (\sum_{i=1}^{n} O_i)^2).\,(n\sum_{i=1}^{n} P_i^2 - (\sum_{i=1}^{n} P_i)^2)}} \quad (11)$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(P_i - \bar{P}_i)^2} \quad (12)$$

Table 5
Algebraic equations generated by multiple linear regression (MLR) and Hargreaves Samani (HS) models, adjusted to estimate ET_0 in the pooled scenario.

Scenario	Algebraic equation MLR	R^2	Adjusted HS	R^2
Pooled	$ET_0 = -3.12^{**} + 0.03771^{ns}\,Tmax + 0.00263^{ns}\,Tmax^2 + 0.958^{***}Tmin - 0.07869^{****}Tmin^2 - 0.01941^{***}\,Tmax\,Tmin + 0.003502^{***}\,Tmax\,(Tmin^2) - 0.00003909^{***}\,(Tmax^2)\,(Tmin^2)$	0.75	$ET_0 = 1.45886^{***} + 0.63267^{***}ET_0^{HS}$	0.49

Tmax – maximum temperature; Tmin – minimum temperature; R^2 – coefficient of determination; ns – not significant by Student's *t* test; and *, **, *** Significant at levels of 0.05, 0.01 and 0.001 by Student's *t* test, respectively.

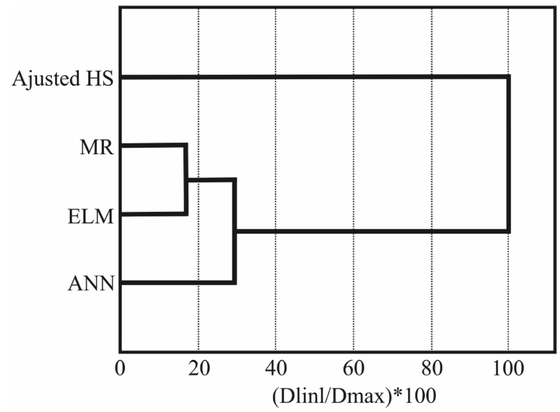

Fig. 8. Hierarchical dendrogram obtained by the Ward algorithm for the MLR, adjusted HS, ANN and ELM models generated with pooled data.

where n is the total number of data tested, and P_i and O_i are the ET_0 values obtained by the FAO-56 PM method and by the other methods (HS, adjusted HS, MLR, ANN and ELM), respectively. RRMSE is dimensionless and presents the perfect fit with a result of 0. The NS and r coefficients are dimensionless and have the perfect fit with a result of 1.

2.9. Multivariate analysis

A multivariate analysis of the data set was performed using cluster analysis (CA). The ET_0 data for the HS, adjusted HS, MLR, ANN, and ELM models were previously standardized to avoid misclassification due to size differences of the data. Subsequently, standardized Euclidean distances as similarity measure were estimated. CA was performed in the data set by the Ward method (Ward, 1963).

2.10. Data processing

Fig. 2 shows the workflow of the ET_0 estimation in this study.

3. Results and discussion

3.1. Local implementation of the models

The MLR, ANN and ELM models, based on air temperature data, presented similar performance criteria (RRMSE, r and NS), with r varying between 0.962 and 0.806 (Fig. 3A), NS ranging from 0.926 to 0.648 (Fig. 3B), and RRMSE ranging from 0.171 to 0.072 (Fig. 3C). The MLR model showed the lowest RRMSE in all cities. In the cities of Espinosa (station 1), Janaúba (station 2), Juramento (station 3) and Montes Claros (station 4), the ANN model had the highest r and NS values, while in the city of Monte Azul (station 5), the r and NS indexes were higher in the ELM model. The MLR, ANN and ELM models, adjusted for the city of Montes Claros (station 4), presented the best performance criteria (RRMSE, r and NS) compared to the other cities (Fig. 3). The algebraic equations that best estimated ET_0 for the MLR and adjusted HS models in the local scenario are presented in Table 3.

The ANN, ELM and MLR models were more accurate than the original and adjusted HS empirical models, according to RRMSE, r and NS performance criteria. The original HS model had higher accuracy in the city of Montes Claros (station 4), with r, NS and RRMSE of 0.881, 0.626 and 0.151, respectively. In contrast, the adjusted HS model showed higher precision in the city of Juramento (station 3), with r, NS and RRMSE of 0.875, 0.763 and 0.104, respectively. Regarding the original HS model, the cities of Janaúba (station 2), Juramento (station 3) and Monte Azul (station 5) presented the least accurate results, with NS of −1.162, RRMSE of 0.240 and r of 0.713, respectively, while the adjusted HS model had lower accuracy in Monte Azul (station 5), with r, NS and RRMSE of 0.710, 0.503 and 0.149, respectively. The use of regression made the ET_0 estimate more accurate by the adjusted HS model compared to the original HS equation (Fig. 3).

Considering the PM model (FAO 56 – PM) as a reference, the original HS model overestimated ET_0 in January, February, March, September, October, November and December and underestimated it in the other months. Regarding the adjusted HS, the model overestimated ET_0 in January, February, November and December and underestimated it in the other months. The behavior presented by the original and adjusted HS models is consistent with the studies reported by Allen et al. (1998), Didari and Ahmadi (2019), and Droogers and Allen (2002), where the HS model also tended to overestimate ET_0 under conditions of high relative humidity. The ET_0 values of the MLR, ANN and ELM models were closer to the ET_0 values estimated by the PM model compared to the adjusted HS and HS models. The MLR, ANN and ELM models showed underestimated and overestimated values ranging from −3.89 to 1.71%, −4.44 and 1.12%, and −3.32 and 1.70%, respectively (Fig. 3).

Table 4 shows the global average performance of the models generated with local data, considering the joint evaluation of the five stations analyzed during the study period. The HS model overestimated ET_0 by 4.8%, while the adjusted HS, MLR, ANN and ELM models underestimated 1.3, 1.1, 1.1 and 1.4%, respectively (Table 4). The ANN models (with r of 0.751 and NS of 0.687) and MLR (with RRMSE of 0.068) presented higher global precision, in addition to providing annual mean ET_0 values closer to those of the reference model, FAO 56 – PM (Table 4). The original HS was the model with the lowest global accuracy, with NS of −0.042 and RRMSE of 0.157; in addition, the annual mean ET_0 estimated by the original HS differed by 87.2 mm from the reference model (FAO 56 – PM), while showing differences between 2.5 and 18.0 mm with respect to the other models (Table 4).

The HS, adjusted HS, MLR, ANN and ELM models were grouped into two clusters (Fig. 5). Cluster 1 consisted of HS and adjusted HS, models that presented the worst performance coefficients (Fig. 3 and Table 4), and ET_0 values farther from the reference model, FAO-56 PM (Fig. 4 and Table 4). Cluster 2 was formed by the MLR, ANN and ELM models (Fig. 5).

The MLR, ANN and ELM models presented the best performance coefficients (Fig. 3 and Table 4) and mean monthly and annual ET_0 values closer to the reference model, FAO-56 PM (Fig. 4 and Table 4), in addition to being grouped in the same Cluster (Fig. 5). In the local scenario, these results indicate the applicability of ANN and ELM artificial intelligence methods and the MLR model to estimate ET_0 using only temperature data.

3.2. Pooled implementation of the models

Similar to the local scenario, the MLR, ANN and ELM models generated with the pooled temperature data from the five stations showed similar performance, with r varying between 0.934 and 0.795 (Fig. 6A), NS ranging from 0.858 to 0.415 (Fig. 6B), and RRMSE ranging from 0.223 to 0.094 (Fig. 6C). The MLR model had better performance in the cities of Espinosa (station 1, with r of 0.900, NS of 0.725 and RRMSE of 0.120), Janaúba (station 2, with r of 0.838 and RRMSE of 0.124) and Montes Claros (station 4, with r of 0.934, NS of 0.858 and RRMSE of 0.096), while the ELM model performed better in the cities of Juramento (station 3, with r of 0.934 and NS of 0.853) and Monte Azul (station 5, with r of 0.811 and NS of 0.616) (Fig. 6). The adjusted HS model had the worst performance in all cities, with r ranging from 0.899 to 0.698 (Fig. 6A), NS ranging from 0.706 to 0.070 (Fig. 6B), and RRMSE ranging from 0.234 to 0.188 (Fig. 6C). The algebraic equations that best estimated ET_0 for the MLR and adjusted HS models, in the pooled data scenario, are presented in Table 5.

Similar to the local scenario, MLR and ELM models underestimated ET_0 during most of the year, showing monthly mean values that were higher than the PM model (FAO 56 – PM) only in March, May and June, when the MLR model was analyzed, and March, when the ELM model was analyzed. The ANN model underestimated ET_0 during the dry and cold period of the year (May to September), while the adjusted HS model overestimated ET_0 in January, February, November, and December. The adjusted HS, MLR, ANN and ELM models presented underestimated and overestimated values ranging from −8.74 to 5.91%, −5.19 and 1.11%, −6.65 and 4.20%, and −7.73 and 1.64%, respectively (Fig. 7).

Table 6 shows the global average performance of the models generated with pooled data, considering the joint evaluation of the five stations analyzed during the study period. The adjusted HS, MLR, ANN and ELM models underestimated ET_0 by 1.1, 1.2 0.2, and 2.7%, respectively (Table 6). The MLR (with r of 0.731 and RRMSE of 0.091) and ELM (with NS of 0.601) models presented the best global performance, although the ANN model provided annual mean ET_0 values closer to the reference model (FAO 56 – PM) (Table 6). The adjusted HS model had the worst global performance, with r of 0.661, NS of 0.305 and RRMSE of 0.171 (Table 6).

The MLR, adjusted HS, ANN and ELM models were grouped into two clusters (Fig. 8). Cluster 1 was formed by the adjusted HS, the model that presented the worst performance coefficients (Fig. 6 and Table 6). Cluster 2 was formed by the MLR, ANN and ELM models (Fig. 8).

Similar to the observations in the local scenario, the MLR, ANN and ELM models had the best performance coefficients (Fig. 6 and Table 6) and were grouped in the same cluster (Fig. 8) in the pooled data scenario. These results indicated the best applicability of ANN and ELM artificial intelligence methods and the MLR model to estimate ET_0, using only temperature data.

The artificial intelligence methods (e.g., ANN and ELM) usually perform better than linear models (e.g., MLR) because they are efficient for modeling phenomena of linear and nonlinear nature, as well as considering possible interactions between variables. The MLR model used in this research presented similar results to the models based on artificial intelligence (ANN and ELM), as observed in other studies (Huo et al., 2012; Ribeiro et al., 2019; Tabari et al., 2012). This can be justified by the fact that the linear, quadratic and interaction effects were considered in some of the MLR models.

The locally trained models were more accurate than the models with pooled temperature data, which is consistent with the data reported by Shiri et al. (2014) in Iran and contrasts with the work of Feng et al. (2017b) in China. However, the generation of models using a combined approach allows the estimation of ET_0 in cases of unavailability of local meteorological data.

The present study confirmed the ability of the ANN, ELM and MLR models to estimate ET_0 only with air temperature data as input. Compared with the ANN model based on MLP, the ELM demonstrated much faster performance when executing the training and testing processes in R software. The MLR models were easier to apply because they provide explicit algebraic equations for the calculation of ET_0 (Table 3 and 5).

Therefore, ANN, ELM and MLR models can be used in irrigation planning and management, thus allowing more precise estimation of crop water requirements through air temperature data; however, the MLR model presents the advantage of the use of explicit algebraic equations, which facilitates their use by farmers and technicians to accurately estimate ET_0.

4. Conclusions

In this study, the applicability of the Hargreaves Samani (HS), adjusted HS, artificial neural network (ANN), multiple linear regression (MLR) and extreme learning machine (ELM) models was evaluated to estimate the reference evapotranspiration (ET_0) based on temperature data from the Verde Grande River basin, southeastern Brazil. This study was carried out in two parts. In the first part (local scenario), all the models were trained, calibrated and validated separately at each station. In the second part (pooled scenario), the meteorological data from all the stations were grouped to train and calibrate the models and then tested separately at each station. To test and validate the models, the Penman-Monteith model (FAO-56 PM) for ET_0 estimation was considered as a reference. The main conclusions can be summarized as follows.

(1) The ANN, MLR and ELM models showed a great ability to estimate ET_0; moreover, they were more accurate than the original and adjusted HS models, both in the local and pooled scenarios.
(2) The ANN, MLR and ELM models had very similar accuracies, but the MLR models were easier to use, as they calculated the ET_0 using explicit algebraic equations.
(3) Although the ANN, MLR and ELM models showed better adjustment in the local scenario, in the pooled scenario, these models allowed the accurate estimation of ET_0 in cases of unavailability of local meteorological data.
(4) The use of the linear regression model to fit the HS equation increased the accuracy of this model to estimate ET_0.

Acknowledgments

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for providing scholarships to the first author and financial support. Espaço da Escrita – Pró-Reitoria de Pesquisa – UNICAMP – for providing the language services.

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 56. Fao, Rome 300 (9), D05109.
Almorox, J., Quej, V.H., Martí, P., 2015. Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. J. Hydrol. 528, 514–522. https://doi.org/10.1016/j.jhydrol.2015.06.057.
Anapalli, S.S., Ahuja, L.R., Gowda, P.H., Ma, L., Marek, G., Evett, S.R., Howell, T.A., 2016. Simulation of crop evapotranspiration and crop coefficients with data in weighing lysimeters. Agric. Water Manag. 177, 274–283. https://doi.org/10.1016/j.agwat.2016.08.009.
Antonopoulos, V.Z., Antonopoulos, A.V., 2017. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. Comput. Electron. Agric. 132, 86–96. https://doi.org/10.1016/j.compag.2016.11.011.
Cao, J., Zhang, K., Luo, M., Yin, C., Lai, X., 2016. Extreme learning machine and adaptive sparse representation for image classification. Neural Networks 81, 91–102. https://doi.org/10.1016/j.neunet.2016.06.001.
Citakoglu, H., Cobaner, M., Haktanir, T., Kisi, O., 2014. Estimation of monthly mean reference evapotranspiration in Turkey. Water Resour. Manag. 28, 99–113. https://doi.org/10.1007/s11269-013-0474-1.
Cobaner, M., 2013. Reference evapotranspiration based on class A pan evaporation via wavelet regression technique. Irrig. Sci. 31, 119–134. https://doi.org/10.1007/s00271-011-0297-x.
Cobaner, M., Citakoglu, H., Haktanir, T., Kisi, O., 2017. Modifying Hargreaves-Samani equation with meteorological variables for estimation of reference evapotranspiration in Turkey. Hydrol. Res. 48, 480–497. https://doi.org/10.2166/nh.2016.217.
Didari, S., Ahmadi, S.H., 2019. Calibration and evaluation of the FAO56-Penman-Monteith, FAO24-radiation, and Priestly-Taylor reference evapotranspiration models using the spatially measured solar radiation across a large arid and semi-arid area in southern Iran. Theor. Appl. Climatol. 136, 441–455. https://doi.org/10.1007/s00704-018-2497-2.
Doorenbos, J., Pruitt, W.O., 1977. Crop water requirements. FAO irrigation and drainage paper 24. Land and Water Development Division, FAO, Rome 144.
Dou, X., Yang, Y., 2018. Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. Comput. Electron. Agric. 148, 95–106. https://doi.org/10.1016/j.compag.2018.03.010.
Droogers, P., Allen, R.G., 2002. Estimating reference evapotranspiration under inaccurate data conditions. Irrig. Drain. Syst. 16, 33–45. https://doi.org/10.1023/A:101550832.
Duan, L., Bao, M., Miao, J., Xu, Y., Chen, J., 2016. Classification based on multilayer extreme learning machine for motor imagery task from EEG signals. Procedia Comput. Sci. 88, 176–184. https://doi.org/10.1016/j.procs.2016.07.422.
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. https://doi.org/10.1016/j.agrformet.2018.08.019.

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

Feng, Y., Peng, Y., Cui, N., Gong, D., Zhang, K., 2017b. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. Comput. Electron. Agric. 136, 71–78. https://doi.org/10.1016/j.compag.2017.01.027.

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

Gocic, M., Petković, D., Shamshirband, S., Kamsin, A., 2016. Comparative analysis of reference evapotranspiration equations modelling by extreme learning machine. Comput. Electron. Agric. 127, 56–63. https://doi.org/10.1016/j.compag.2016.05.017.

Hargreaves, G.H., Samani, Z.A., 1985. Reference Crop Evapotranspiration from Temperature. Appl. Eng. Agric. 1 (2), 96–99. https://doi.org/10.13031/2013.26773.

Huang, G. Bin, Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70, 489–501. https://doi.org/10.1016/j.neucom.2005.12.126.

Huo, Z., Feng, S., Kang, S., Dai, X., 2012. Artificial neural network models for reference evapotranspiration in an arid area of northwest China. J. Arid Environ. 82, 81–90. https://doi.org/10.1016/j.jaridenv.2012.01.016.

INMET (2019). Certification awarded to Instituto Nacional de Meteorologia (INMET) on 01 March 2012: the development, collection and supply of data and meteorological products and services is ISO 9001:2008 certified, accessed 28 Jun 2019, < http://www.inmet.gov.br/portal/index.php?r=home/page&page=qualidade > .

Kisi, O., Cimen, M., 2009. Evapotranspiration modelling using support vector machines. Hydrol. Sci. J. 54, 918–928. https://doi.org/10.1623/hysj.54.5.918.

Kisi, O., Sanikhani, H., 2015. Modelling long-term monthly temperatures by several data-driven methods using geographical inputs. Int. J. Climatol. 35, 3834–3846. https://doi.org/10.1002/joc.4249.

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Niazi, F., 2015. Long-term monthly evapotranspiration modeling by several data-driven methods without climatic data. Comput. Electron. Agric. 115, 66–77. https://doi.org/10.1016/j.compag.2015.04.015.

Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W., Pruitt, W.O., 2002. Estimating evapotranspiration using artificial neural network. J. Irrig. Drain. Eng. 128 (4), 224–233. https://doi.org/10.1061/(ASCE)0733-9437(2002)128:4(224).

Ribeiro, A.F.S., Russo, A., Gouveia, C.M., Páscoa, P., 2019. Modelling drought-related yield losses in Iberia using remote sensing and multiscalar índices. Theor. Appl. Climatol. 136, 203–220. https://doi.org/10.1007/s00704-018-2478-5.

Shiri, J., Nazemi, A.H., Sadraddini, A.A., Landeras, 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. https://doi.org/10.1016/j.compag.2014.08.007.

Smith, M., 2000. The application of climatic data for planning and management of sustainable rainfed and irrigated crop production. Agric. For. Meteorol. 103, 99–108. https://doi.org/10.1016/S0168-1923(00)00121-0.

Smith, M., Allen, R.G., Monteith, J.L., Perrier, A., Pereira, L.S., Segeren, A., 1991. Expert consultation on procedures for revision of FAO guidelines for prediction of crop water requirements. FAO, Rome.

Tabari, H., Kisi, O., Ezani, A., Talaee, P.H., 2012. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. J. Hydrol. 444–445, 78–89. https://doi.org/10.1016/j.jhydrol.2012.04.007.

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

Traore, S., Wang, Y.M., Kerh, T., 2010. Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. Agric. Water Manag. 97, 707–714. https://doi.org/10.1016/j.agwat.2010.01.002.

Valayamkunnath, P., Sridhar, V., Zhao, W., Allen, R.G., 2018. Intercomparison of surface energy fluxes, soil moisture, and evapotranspiration from eddy covariance, large-aperture scintillometer, and modeling across three ecosystems in a semiarid climate. Agric. For. Meteorol. 248, 22–47. https://doi.org/10.1016/j.agrformet.2017.08.025.

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244. https://doi.org/10.1080/01621459.1969.10500963.

Xu, G., Xue, X., Wang, P., Yang, Z., Yuan, W., Liu, X., Lou, C., 2018. A lysimeter study for the effects of different canopy sizes on evapotranspiration and crop coefficient of summer maize. Agric. Water Manag. 208, 1–6. https://doi.org/10.1016/j.agwat.2018.04.040.

Yan, C., Zhao, W., Wang, Y., Yang, Q., Zhang, Q., Qiu, G.Y., 2017. Effects of forest evapotranspiration on soil water budget and energy flux partitioning in a subalpine valley of China. Agric. For. Meteorol. 246, 207–217. https://doi.org/10.1016/j.agrformet.2017.07.002.

Yin, Y., Wu, S., Zheng, D., Yang, Q., 2008. Radiation calibration of FAO56 Penman-Monteith model to estimate reference crop evapotranspiration in China. Agric. Water Manag. 95, 77–84. https://doi.org/10.1016/j.agwat.2007.09.002.