Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil

Maurício Bruno Prado da Silva[a], João Francisco Escobedo[a,*], Taiza Juliana Rossi[a], Cícero Manoel dos Santos[b], Sílvia Helena Modenese Gorla da Silva[c]

[a] Department of Rural Engineering - FCA, UNESP, Botucatu, SP, Brazil
[b] Agriculture College - UFPA, Altamira, PA, Brazil
[c] Experimental Campus – UNESP, Registro, SP, Brazil

ARTICLE INFO

ABSTRACT

This study describes the comparative study of different methods for estimating daily global solar irradiation (H): Angstrom-Prescott (A-P) model and two Machine Learning techniques (ML) – Support Vector Machine (SVM) and Artificial Neural Network (ANN). The H database was measured from 1996 to 2011 in Botucatu/SP/Brazil. Different combinations of input variables were adopted. *MBE, RMSE, d* Willmott, *r* and r^2 statistical indicators obtained in the validation of A-P and SVM and ANN models showed that: SVM technique has better performance in estimating H than A-P and ANN models. A-P model has better performance in estimating H than ANN.

1. Introduction

Knowledge of global solar irradiation (H) is of utmost importance in climate studies on renewable energy, architecture projects and agriculture (growth models and yield of agricultural crops and evapotranspiration estimates) (Chen et al., 2004; Souza et al., 2005; Almorox et al., 2005; Hsiao et al., 2008; Bosch et al., 2008; El-Sebaii et al., 2009). Brazil is a country of great continental dimensions and knowledge of annual solarimetric availability for different applications, mainly projects of power generation and cogeneration is also very important (Martins et al., 2012; Castillo et al., 2016). However, Brazil has a tremendous lack of solarimetric information caused by high cost and maintenance of solarimetric stations. Therefore, many researchers have been developing and adjusting models that allow estimating different types of solar radiation such as: global, diffuse, direct and spectral radiation (UV, Photosynthetically Active Radiation – PAR and Near Infrared Radiation – NIR). In general, these models are categorized into different classes: statistical, empirical, physical (radiative transfer) and more recently Machine Learning techniques (ML) (Oliveira et al., 2002; Soares et al., 2004; Jiang, 2008; Martins et al., 2008; Escobedo et al., 2012; Santos et al., 2014, 2016; Lyra et al., 2015).

Statistical models are usually more recommended because of their simplicity of use, as they need other input variable for routine measurements in meteorological stations. Their disadvantage is that they are only valid for locations in which they are generated, adjusted or for regions of similar climate. The main and most used statistical model is the Angstrom-Prescott model (A-P), which estimates H in daily or monthly partition from daily sunshine duration values (n). Hargreaves and Samani (1982) and Bristow and Campbell (1984) are the two other models which use air temperature variation as an input variable. There are also other modified models from different locations which associate air temperature with other input variables (Hunt et al., 1998; Bechini et al., 2000; Almorox et al., 2011a; Bojanowski et al., 2013).

Radiative transfer models are more complex than statistical models and require mixed input meteorological parameters measured by satellites on the Earth's surface and atmosphere (Madkour et al., 2006). The disadvantage of these models is the need for detailed information of difficult access on ozone, aerosols, gases and water vapor concentrations in the atmosphere (Gueymard, 2003; Kaushika et al., 2014). Radiative transfer models require extensive computational work on complex atmospheric transfer processes as a result of temporal and spatial variations of aerosols and water vapor concentrations (Dai and Fang, 2014; Zhang et al., 2014).

ML models are able to solve complex problems and have been successfully applied to estimate solar radiation (Rehman and Mohandes, 2008). The technique allows modeling a system in which

only the input and output variables are known. The main advantage of using ML models is their capacity of generalizing and optimizing time (Oliveira et al., 2006). The main ML models used were as follows: Adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN), Genetic Programming (GP) and Support Vector Machine (SVM). Some recent studies have concluded that the SVM technique has better performance in estimating solar radiation than ANN, ANFIS techniques and other numerical methods. In particular, *H* estimates by SVM and ANN models have been studied and compared with statistical models, and the results have shown that the perfor-mance of ML (SVM and ANN) models are similar in some cases and better in other cases than statistical models (Elizondo et al., 1994; Tymvios et al., 2005; Chen et al., 2013; Piri et al., 2015; Quej et al., 2017). The SVM technique shows better performance than the A-P model in several cases, because the formulation of SVM involves the minimization concept of the structural risk, as an opposing approach to minimization of the empirical risk, which is widely used in the methods of statistical learning. The minimization of the structural risk attenu-ates the upper limit in the generalization error in contrast to the minimization of the empirical risk, which makes the error minimum in data of training. That difference is the reason for the SVM's better generalization potential (Shamshirband et al., 2014; Ramedani et al., 2014).

In southeastern Brazil, there are few studies on *H* modeling using ML, so adjustment and validation of ML models in estimating *H* is of great interest for mapping and potential using of solar radiation in projects in Brazil (Lima et al., 2016).

The present study presents a comparative study between the A-P statistical model and ML models. A 16-year daily global solar irradia-tion database from 1996 to 2011 was used to generate and validate A-P models and two SVM and ANN techniques. For SVM and ANN techniques, four different kinds of architecture combining input variables in the model were studied: model 1 of SVM and ANN has as input variables fractional daily sunshine duration ($r'=n/N$, where n is daily sunshine duration and N is day length) and solar radiation at the top of the atmosphere (H_O) equal to the A-P model; and in the following models 2, 3 and 4, other variables were added, one by one, such as air temperature, precipitation and relative humidity, respec-tively. The results of validation, comparison between measurements and model estimates by correlation (r) and statistical indicators (*MBE*, *RMSE* and *d* Willmott), allowed classifying the performance of models in *H* estimates and expansion of climate series.

2. Material and method

2.1. Study location and data

Data used in this study were measured in the Solar Radiometric Station at the College of Agricultural Sciences (FCA), University of the State of São Paulo (UNESP), located in Botucatu (22°53′S latitude, 48°26′W longitude and 786 m altitude). Botucatu is a municipality located in the midwestern region of São Paulo state, 1,482,642 km² total area. The city has high altitude gradient between 400 and 500 m in the lowest region and between 700 and 900 m in the mountainous region. That difference causes changes in air temperature and winds. With Cerrado and Atlantic forest biomes, and according to the Köppen climatic classification, the climate of the Botucatu region is **Cwa**, characterized as altitude tropical climate, hot and humid summer with high precipitation, dry winter and average temperature of the warmest month higher than 22 °C. Topography and climate in the region are very favorable for agriculture and solarimetric projects. Moreover, high prevalence of sugarcane and eucalyptus crops in the region is observed (Santos and Escobedo, 2016).

Data on daily sunshine duration, air temperature (maximum and minimum), precipitation and air relative humidity for the 1996−2011 period were used. Global solar irradiance (I, W m⁻²) was monitored

using an Eppley PSP pyranometer with 4.1% error (Reda et al., 2008). In the acquisition of I data, a Campbell Scientific CR23X datalogger was used, operating at frequency of 1 Hz and storing averages every 5 min. Those data have been subjected to rigorous quality control (for elimination of spurious or inconsistent values) through programs developed for calculating integrated irradiation on the day *H* (Chaves and Escobedo, 2000). Daily sunshine duration data (*n*, hours) were obtained by a Campbell-Stokes sunshine recorder; precipitation (P, mm) measured by an Ota Keiki Seisakusho rain gauge; maximum and minimum air temperature (T, °C) measured by mercury and alcohol bulb thermometer, respectively; and air relative humidity (RH, %) measured by a hygrometer according to the World Meteorological Organization, WMO (1981).

2.2. Angstrom-Prescott Model (A-P)

Several models have been suggested to estimate *H* using daily sunshine duration as input variable. The most widely known model to estimate *H* was proposed by Angström (1924) and later modified by Prescott (1940) using Eq. (1): wherein (H_O) is the solar radiation at the top of the atmosphere, (*n*) is the daily sunshine duration and (*N*) is the day length:

$$\frac{H}{H_o} = a + b \times \left(\frac{n}{N}\right) \tag{1}$$

Coefficient "*a*" can be interpreted as the *H* fraction that reaches the Earth's surface on a cloudy day, being dependent on the type and thickness of clouds. Coefficient "*b*" is a supplement which gives the total *H*. The sum ($a+b$) is the potential fraction of solar radiation at the top of the atmosphere available to reach the surface (i.e., *H* on a clear sky day). *Ho* depends on the latitude, solar declination, timing angle and day length (*N*) (Souza et al., 2016). Coefficients "*a*" and "*b*" of the A-P model were determined by the least square method. This method minimizes mean square error of the estimated data compared to measured data.

2.3. Support Vector Machine (SVM) with the Sequential Minimal Optimization (SMO) algorithm

Support Vector Machine (SVM) is a supervised learning technique based on the statistical learning theory (Vapnik, 1995). More detailed information on SVM can be found at Vapnik (1998). Due to its ability to provide excellent generalization performance, SVM has become a powerful tool for resolving problems of pattern recognition, classifica-tion, prediction, and regression (Shevade et al., 2000). Solution for regression problems using SVM can be given through an iterative algorithm called Sequential Minimal Optimization (SMO) (Smola and Schölkopf, 1998). Subsequently, improvements in SMO have been suggested (Shevade et al., 2000). SMO is a simple algorithm that quickly solves the problem of the lowest possible optimization with two Lagrange multipliers (Platt, 1998; Smola and Schölkopf, 2004). Computational speed and ease of implementation are favorable char-acteristics in the use of the SMO algorithm. The advantage of its use over other techniques is because it is based on the structural risk minimization principle, which attempts to minimize a generalization error upper limit instead of minimizing the local training error (Chen et al., 2013). In addition, it offers a unique solution and estimates regression using a set of kernel functions which are defined in a high-dimensional space, making data linearly separable. The used Kernel functions are generally polynomial, sigmoidal and the radial basis function (RBF). In this study, RBF is used for regression because of computational efficiency, simplicity and adaptation for optimization of complex problems.

When using RBF, it is necessary to properly adjust C (cost), γ (gamma) and ε (epsilon) parameters. C, γ and ε values are tested and
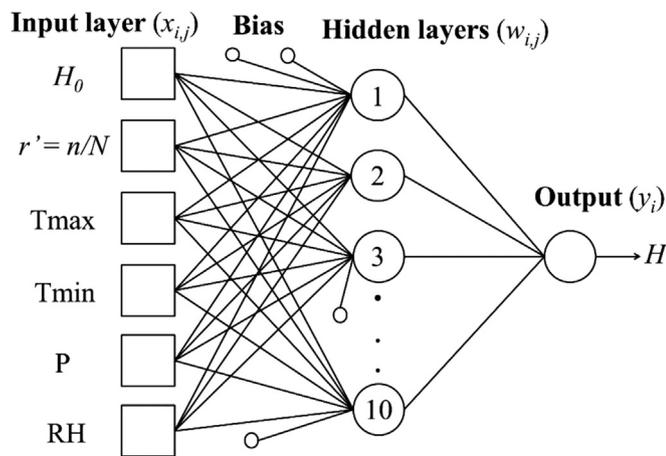
Fig. 1. Block diagram showing the Artificial Neural Network (ANN) architecture used, adapted from Haykin (1998).

Fig. 2. Comparison between the average inter-year H value with standard deviation and the average irradiance value of the year for the month of January in the series.

those with the best accuracy in the cross validation are chosen. After several parameter tests (C = 100, γ = 0.3 and ε = 0.001), values equal to those obtained by Ramedani et al. (2014) and Mohammadi et al. (2015a) were used.

2.4. Multilayer Perceptron (MLP) with Back Propagation algorithm (BP)

Multilayer Perceptron (MLP) was the ANN used. MLP maps sets of input data for a set of output data. This technique is widely used in modeling to solve complex problems. Fig. 1 shows the MLP structure, wherein the first layer is the input ($x_{i,j}$), the second has one or more hidden layers of compute nodes with connection weights ($w_{i,j}$) and the third layer corresponds to the output of compute nodes (y_i), (Lyra et al., 2015). Input signals are sent to the hidden layer. Then, hidden and output layers multiply input signals by a set of weights.

The typical MLP with a hidden layer can be modeled as Eq. (2) (Lam et al., 2008b):

$$y_i = \sum_{j=1}^{n} w_{i,j} x_{i,j} + \theta_i \qquad (2)$$

Where θ_i is the bias of neuron i. Each entry is multiplied by a connection weight. Output of neurons is calculated by applying a nonlinear activation function, Eq. (3), which is typically standard sigmoid (Rehman and Mohandes, 2008).

$$f(x) = sigmoid(x) = \frac{1}{(1 + exp(-x))} \qquad (3)$$

There are many specific learning algorithms for certain neural network models. MLP was trained using supervised learning algorithm Back Propagation (BP) and the momentum term. In this algorithm, the value of each output layer is used to update the weight of the previous layer. BP interactively learns the joint processing of data training examples. Weight adjustment in the iteration depends on the learning rate and momentum. The learning rate for each interaction controls the size of weight changes and bias.

2.5. Software used

Waikato Environment for Knowledge Analysis (WEKA) toolbox was used to train and validate H data with the SMO algorithm for SVM and BP for ANN. WEKA is a set of ML algorithms containing tools for data preprocessing, classification, regression and association and visualization rules (Witten et al., 2011). SMO is used with the RBF Kernel function for model formation. In ANN models with BP algorithm, the following values were considered: learning rate = 0.3; momentum=0.2 and number of iterations=500. The hidden layers were tested, ranging
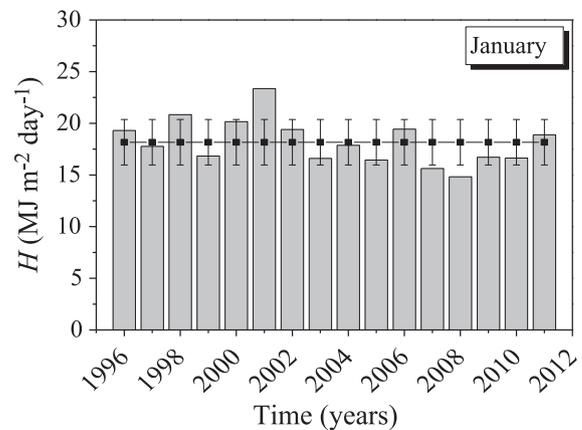
from 1 to 10, but the WEKA default value was adopted for the best fit found. In WEKA, the pattern of hidden layers is defined as "α"= [(input variables+classes)/2]. The sigmoidal activation function is adopted.

2.6. Validation database of models: selection of atypical and typical years

In the validation of models generated in this work, a 2-year database of H was used, called typical and atypical year, separated from the total 16-year database 1996–2011. The selection of typical and atypical years was accomplished through statistical analysis of data, where the average inter-year H value with standard deviation of each month and the average irradiance value of each month per year were compared in each month of the year. Example of application of this methodology can be found in Fig. 2, where the horizontal line with dumbbells represents the average inter-year irradiation with standard deviation, and columns represent the average irradiation of each year for the month of January.

The typical year selection criterion is similar to the selection process of the typical meteorological year published by the World Meteorological Organization (WMO) in 1981. For the month of January, the typical year was 2004 – column closest to the average inter-year irradiation, while 2001 was atypical – column farthest from the average inter-year irradiation. The results found for the remaining months are shown in Table 1, which shows the monthly constitution of typical and atypical years of the total 16- year database.

The organization structure of input data begins with the initial columns corresponding to the input variables and the last column to the output variable, columns separated by comma. The input files in the .dat format for training and validation of SVM and ANN were transformed into .arff file, which is the extent necessary for processing data in the WEKA toolbox.

2.7. Statistical indicators

There are several statistical indexes used to evaluate model performances. For solar radiation models the following indexes are usually used: Mean Bias Error (*MBE*), Relative Mean Bias Error (*rMBE*), Root Mean Square Error (*RMSE*), Relative Root Mean Square Error (*rRMSE*), correlation coefficient (*r*), determination coefficient (r^2) and the Willmott Concordance Index (*d*) (Santos et al., 2016). The equations below were used to evaluate the performance of the generated models:

$$MBE = \frac{\sum_{i=1}^{M} (Y_i - X_i)}{M} \qquad (4)$$

Table 1
Typical and atypical year obtained from the 16-year database.

		Month											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Year	Typical	2004	2000	2008	2005	2005	2005	2002	2003	2000	2003	2003	2005
	Atypical	2001	2005	2011	2008	2003	1997	2009	1998	2009	2001	1998	2011

$$rMBE\,(\%) = 100 \times \frac{\left(\frac{\sum_{i=1}^{M}(Y_i - X_i)}{M}\right)}{\overline{X}} \tag{5}$$

$$RMSE = \left[\frac{\sum_{i=1}^{M}(Y_i - X_i)^2}{M}\right]^{\frac{1}{2}} \tag{6}$$

$$rRMSE\,(\%) = 100 \times \left[\frac{\sum_{i=1}^{M}(Y_i - X_i)^2}{\overline{X}}\right]^{\frac{1}{2}} \tag{7}$$

$$r = \frac{\sum_{i=1}^{M}(Y_i - \overline{Y_i})(X_i - \overline{X_i})}{\sqrt{\sum_{1}^{M}(Y_i - \overline{Y_i})\sum_{i=1}^{M}(X_i - \overline{X_i})}} \tag{8}$$

$$r^2 = \frac{\left[\sum_{i=1}^{M}(Y_i - \overline{Y_i})(X_i - \overline{X_i})\right]^2}{\sum_{1}^{M}(Y_i - \overline{Y_i})\sum_{i=1}^{M}(X_i - \overline{X_i})} \tag{9}$$

$$d = 1 - \frac{\sum_{i=1}^{M}(Y_i - X_i)^2}{\sum_{i=1}^{M}(|Y_i - \overline{X}| + |X_i - \overline{X_i}|)} \tag{10}$$

where Y_i represents the estimated values, $\overline{Y_i}$ are the average estimated values, X_i are the measured values, $\overline{X}$ are the average measured values and M is the number of observations.

MBE (or *rMBE*) provides information on the performance of long-term models, allowing a comparison of the actual deviation between estimates and measures. The ideal *MBE* value is "zero". The disadvantage of this method is that an overestimation cancels an underestimation. *RMSE* (or *rRMSE*) provides information on the short-term performance. The *RMSE* value is always positive, and the lower the *RMSE* values obtained, the better the model performance. The adjustment index "*d*" ranging from 0 to 1 represents total maladjustment and adjustment, respectively, between estimates and measures (Willmott, 1981). A classification scale for different *rRMSE* intervals to evaluate the performance of models is used (Jamieson et al., 1991; Li et al., 2013): excellent if *rRMSE* < 10%; good if 10%≤*rRMSE* < 20%; acceptable if *rRMSE*≤20% < 30% and poor if *rRMSE*≥30.

3. Results and discussion

3.1. Correlation between atmospheric transmissivity of global solar irradiation (H/H_O) and fractional daily sunshine duration (n/N): Angstrom-Prescott Model (A-P)

Fig. 3 shows the correlation between atmospheric transmissivity of global solar irradiance ($Kt = H/H_O$) and fractional daily sunshine duration (n/N) for the database measured in the 1996–2011 period in Botucatu, SP. Overall, 5685 days were used, of which 921 days were for $Kt \leq 0.35$ (cloudy sky condition), and 3387 days were for $0.35 < Kt \leq 0.65$ (partly cloudy sky condition) and 1377 days with $Kt > 0.65$ (clear sky condition). The correlation is linear throughout the variation range n/N between 0 and 1. Correlation scattering is similar to that from most studies using the Angstrom-Prescott equation (Bakirci, 2009).

Correlation scattering is high, for each n/N value, there is a wide range of variation in H/H_O values. The effect is because of the great variability of cloud concentrations (when combining layer type, num-
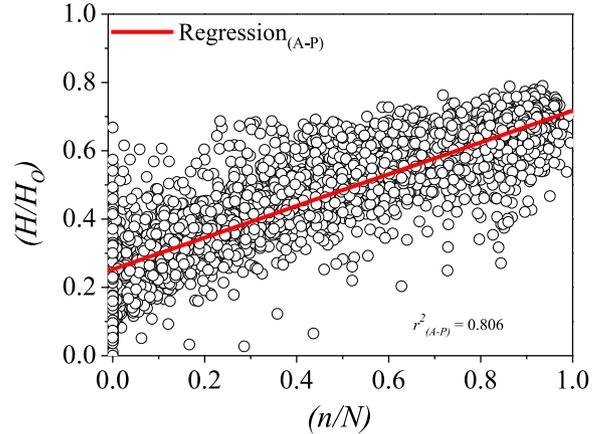
Fig. 3. Correlation between atmospheric transmissivity of global solar irradiation (H/H_O) and fractional daily sunshine duration (n/N); the red color line is the linear regression line in the 1996–2011 period. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ber and thickness), water vapor and aerosols in the atmosphere that absorb and spread global radiation differently in different types of sky coverage for the same n/N values. There are many cloud cover combinations that will generate the same n/N value, but one cannot expect that each combination produces a single H/H_O value, as can be seen in n/N values close to zero. The Eq. (11) of A-P obtained in the correlation of Fig. 3 by linear regression is:

$$(H/H_O) = 0.253 + 0.465 \times (n/N) \tag{11}$$

The determination coefficient (r^2) is 0.806 and the value of the minimum H fraction a is 0.253, which provides total H ($b = 0.465$) or the maximum atmospheric fraction ($a + b = 0.718$) are similar to values obtained in several locations in Brazil and are within the variation range of coefficients ($a = 0.265 \pm 0.052$) ($a + b = 0.689 \pm 0.058$) and r^2, whose average is 0.778 ± 0.095 (Table 2).

Despite the wide use of A-P model, much because of its practicality and not much because of its accuracy; it depends on the climate in the location of the measures. The permanent adjustment of the equation coefficients a and b must be assessed periodically because of climatic variations. Other limitation of the method is the sunshine measure, which is not accurate, mainly in tropical regions where air relative humidity is high, making it difficult burning the heliograph tape.

Climatic conditions, cloud type and thickness, water vapor and aerosol concentrations in the atmosphere are factors responsible for variations of "*a*" and "*a+b*" coefficients in each location. The "*a*" and "*b*" values are found in this study close to those suggested by Allen et al. (1998) by the FAO-56 Bulletin, which is $a = 0.25$ and $b = 0.50$ to be a universal equation that meets locations where there are no solar global radiation measures and no calibration has been carried out for improved a and b values. Therefore, the coefficients obtained can be used in locations with the same climatic conditions.

3.2. Validation of the Angstrom-Prescott model (A-P)

Fig. 4(a, b) shows the correlations obtained in the validation between estimates and H measures for typical and atypical databases.

Table 2

Angstrom-Prescott coefficients (A-P) obtained in Brazil by several authors.

Authors	Cities	Coefficients			
		a	*b*	*a+b*	*r²*
Mota et al. (1977)	55 locations (Brazil)	0.170–0.31	0.41–0.57	0.71–0.74	–
Azevedo et al. (1981)	Fortaleza (CE)	0.27	0.36	0.63	0.74
Araújo et al. (2001)	São Paulo (SP)	0.29	0.36	0.65	–
Nicácio et al. (2001)	Maceió (AL)	0.32	0.37	0.69	0.76
Blanco and Sentelhas (2002)	Piracicaba (SP)	0.23	0.5	0.73	0.84
Pacheco and Bastos (2002)	Capitão Poço (PA)	0.30	0.34	0.64	0.70
Texeira et al. (2002)	Juazeiro (BA)	0.26	0.32	0.58	0.81
Dantas et al. (2003)	Lavras (MG)	0.23	0.49	0.72	0.79
Santos et al. (2003)	Ilha Solteira (SP)	0.26	0.47	0.73	0.81
Dallacort et al. (2004)	Palotina (PR)	0.21	0.39	0.6	0.84
Back (2005)	Urussanga (SC)	0.23	0.49	0.71	0.72
Dornelas et al. (2006)	Brasília (DF)	0.28	0.49	0.77	0.81
Pilau et al. (2007)	Araras (SP)	0.24	0.46	0.7	0.72
Pereira et al. (2010)	Pedra Azul (MG)	0.27	0.33	0.59	0.57
Torres et al. (2010)	Canavieiras (BA)	0.37	0.27	0.63	0.99
Carvalho et al. (2011)	Seropédica (RJ)	0.28	0.43	0.72	0.82
Andrade Júnior et al. (2012)	Parnaíba (PI)	0.32	0.46	0.78	0.63
Belúcio et al. (2014)	Macapá (AP)	0.27	0.43	0.7	0.82
Martim et al. (2014)	Sinop (MT)	0.27	0.47	0.74	0.86
Berusky et al. (2015)	Ponta Grossa (PR)	0.14	0.48	0.62	–
Souza et al. (2016)	3 locations (AL)	0.24–0.34	0.38–0.48	0.71–0.73	–

Abbreviations represent the following Brazilian states: AP = Amapá, AL = Alagoas, BA = Bahia, CE = Ceará, DF = Federal District, MT = Mato Grosso, MG = Minas Gerais, PA = Pará, PI=Piauí, RJ = Rio de Janeiro, SP = São Paulo, PR = Paraná, SC = Santa Catarina.
– Information not provided by the authors.

Distribution of values estimated by the A-P equation and measures in both validation conditions are in linear correlation with the ideal lines of comparison (1:1). Linear regression equations with correlation coefficients r_t = 0.942 ("t" index indicates typical year) and r_a = 0.939 ("a" index indicates atypical year) show that the A-P equation can estimate H with determination coefficients r_t^2 = 0.887 (Fig. 4a) and r_a^2 = 0.882 (Fig. 4b). Linear correlation coefficients (r) obtained in this study are higher than the r value of = 0.939 and is at the same order of magnitude as r = 0.89–0.98 determined by Li et al. (2011) (Table 3).

Table 3 shows statistical indexes ($rMBE$, $RMSE$, r) obtained in the model validation and those from other locations. Values of $rMBE^t$ $_{(A-P)}$ = −3.0% and $rMBE^a_{(A-P)}$ = 1.1% in Botucatu are at the same order of magnitude as the experimental error for global radiation measured by the sensor, which is 4.1% (Reda et al., 2008). They are higher than those found by Martim et al. (2014), Sabzipavar et al. (2013), Zhao et al. (2013) and Iziomon and Mayer (2001). Moreover, the results were lower than values found by Trnka et al. (2005), Lam et al. (2008a), Wan et al. (2008), Li et al. (2011) and Berusky et al. (2015). The results obtained by Manzano et al. (2015) alternate values above and below the results from Botucatu. These differences result from adjustments of the model to climate conditions of each location.

Scattering values ($rRMSE^t_{A-P}$ = 13.1% and $rRMSE^a_{A-P}$ = 15.7%) were also considered good results according to the criterion by Jamieson et al. (1991). Local $rRMSE$ values were higher than those found by Iziomon and Mayer (2001), Nicácio et al. (2001), Trnka et al. (2005), Torres et al. (2010), Li et al. (2011), Andrade Júnior et al. (2012), Martim et al. (2014) and Berusky et al. (2015). Local $rRMSE$ values alternate with values obtained by: Lam et al. (2008a), Wan et al. (2008), Liu et al. (2009), Li et al. (2012), Chen et al. (2013), Sabzipavar et al. (2013), Zhao et al. (2013), Mohammadi et al. (2015a), Manzano et al. (2015) and Park et al. (2015). Local $rRMSE$ values were lower than those found by Pereira (2010). Concordance index d^t_{A-P}=0.963 and d^a_{A-P} = 0.959 close to 1 shows good agreement between estimates and measures.

3.3. Training and validation of models of SVM and ANN techniques

The same data base used for generation of the A-P model (Fig. 3) was used in the training of SVM and ANN techniques. The Angstrom-Prescott model and SVM1 and ANN1 models (combination 1) use the
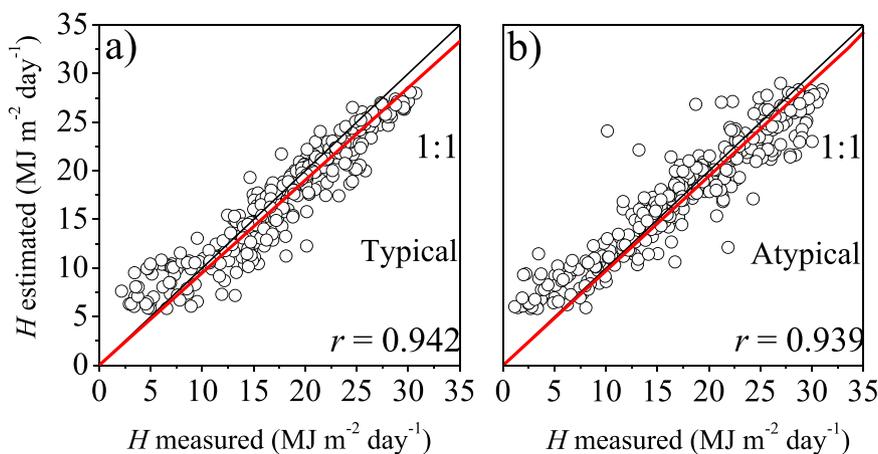
Fig. 4. (a, b). Comparison between values estimated by the method A-P and those measured for (a) typical and (b) atypical years.

Table 3
Statistical indicators obtained for several locations.

Authors	Locations (N°)	Latitudes	rMBE(%)	MBE(MJ m^{-2} day^{-1})	rRMSE(%)	RMSE(MJ m^{-2} day^{-1})	r
Iziomon and Mayer (2001)	2 (Germany)	47°52′ – 47°54′ N	–	(–0.02)–0	–	(0.03–0.7)	–
Nicácio et al. (2001)	Maceió (Brazil)	9°35′ S	–	0.03	–	1.73	0.87
Trnka et al. (2005)	10 (Austria/CzechRepublic)	46°58′ – 50°11′ N	(–4.14)–3.05	(–0.50)–0.32	13.0–17.34	1.41–1.80	–
Lam et al. (2008a)	40 (China)	23°23′ – 49°13′ N	(–14.90)–18.7	(–2.88)–1.84	7.60–32.60	1.19–3.61	–
Liu et al. (2009)	31 (China)	31°09′ – 43°39′ N	–	–	–	1.39–3.08	–
Wan et al. (2008)	41 (China)	22°18′ – 49°13′ N	(–16.20)–18.80	(–3.00)–2.04	7.40–31.30	1.22–3.74	–
Li et al. (2011)	4 (China)	29°40′ – 32°30′ N	–	0.57–1.25	–	1.1–1.64	0.89–0.98
Benghanem and Mellit (2010)	Al-Madinah (Saudi Arabia)	24°33′ N	–	–	–	0.0002	0.97
Pereira et al. (2010)	Pedra Azul (Brazil)	15°14′40′′ S	–	–	–	2.83	–
Torres et al. (2010)	Canavieiras (Brazil)	15°40′ S	–	–	–	1.72	–
Andrade Júnior et al. (2012)	Parnaíba (Brazil)	03°05′ S	–	–	–	0.08	–
Li et al. (2012)	15 (China)	26°34′48′′ – 32°0′ N	–	–	–	1.81–3.39	–
Chen et al. (2013)	3 (China)	38°54′ – 41°44′ N	–	–	14.90–19.40	1.99–2.28	–
Sabzipavar et al. (2013)	15 (Iran)	28°58′48′′ – 38°04′48′′ N	–	(–0.15)–0.07	–	1.66–3.25	–
Zhao et al. (2013)	9 (China)	23°07′ – 45°45′ N	–	(–0.04)–0.09	–	1.72–5.24	–
Martim et al. (2014)	Sinop (Brazil)	11°58′48′′ S	–	–0.02	–	1.85	–
Berusky et al. (2015)	Ponta Grossa (Brazil)	25°05′12′′ S	–	1.58	–	1.64	–
Manzano et al. (2015)	25 (Spain)	36°30′ – 43°29′24′′ N	–	(–1.32)–0.30	–	1.11–3.37	–
Mohammadi et al. (2015a)	Isfahan (Iran)	32°39′41′′ N	–	–	13.9	2.67	–
Park et al. (2015)	22 (South Korea)	33°30′34′′ – 37°52′29′′ N	–	–	–	0.47–2.73	–
Souza et al. (2016)	3 (Brazil)	9°15′ – 9°44′24′′ S	–	(–0.91)–0.22	–	1.95–2.98	–
Present study (2016)[a]	*Botucatu (Brazil)*	*22°53′ S*	*–3.0*	*–0.52*	*13.1*	*2.28*	*0.94*
Present study (2016)[b]	*Botucatu (Brazil)*	*//*	*1.1*	*0.19*	*15.7*	*2.69*	*0.94*

N° = Number of location; – Not informed by the authors; // Repetition of term.

[a] Typical year.

[b] Atypical year.

Table 4
Input variables for *SVM* and *ANN* combinations.

Model	SVM	ANN	Input variable
combination 1	SVM1	ANN1	H_O, n/N
combination 2	SVM2	ANN2	H_O, Tmax, Tmin, n/N
combination 3	SVM3	ANN3	H_O, Tmax, Tmin, P, n/N
combination 4	SVM4	ANN4	H_O, Tmax, Tmin, P, RH, n/N

H_O = solar radiation at the top of the atmosphere, n/N = fractional daily sunshine, Tmax = maximum air temperature, Tmin=minimum air temperature, P = Precipitation and RH=relative humidity.

same input variables H_O, N and n (measured). In addition to combination 1, three new combinations were trained (SVM2, SVM3 and SVM4; ANN2, ANN3 and ANN4), maintaining the same variables of combination 1 and adding the following input variables: maximum and minimum air temperature (T, ºC), precipitation (P, mm) and relative humidity (RH, %), as shown in Table 4. Input data for SVM and ANN models consist of independent and dependent variables. Input variables were selected because of their correlation with H and being more easily monitored and available in stations (Liu and Scott, 2001; Podestá et al., 2004; Behrang et al., 2010).

Fig. 5 shows dispersion between H values estimated by SVM and ANN models and measures, and the straight lines obtained by linear regression and correlation coefficients (r) for typical (r_t) and atypical years (r_a), respectively, for the four combinations. The introduction of a new meteorological variable in each combination modifies the H estimated value, altering scattering, and consequently, changing correlations in both typical and atypical validation bases.

Visually, the dispersion differences among combinations 1, 2, 3 and 4 can be better observed by comparing dispersion in Fig. 5(a, c, e, g) for typical basis and Fig. 5(b, d, f, h) for atypical basis.

In combination 1, values estimated by SVM1 and ANN1 are in linear accordance with the measured values Fig. 5(a, b): correlations (r) are very close to the ideal straight line of comparison (1:1). For SVM1, $r_t = 0.962$ and $r_a = 0.947$ values, while for ANN1 $r_t = 0.924$ and $r_a = 0.932$ values show that H measures and estimates are statistically well correlated. The comparison of (r) values of the training validation of ML techniques with A-P^t model shows that SVM1 has better performance than A-P^t and ANN1 models in both validation conditions. In turn, A-P^t model has better performance than ANN1 in both validation conditions.

In combination 2, SVM2 and ANN2 Fig. 5(c, d) also show correlations (r) very close to the ideal straight line (1:1), ($r_t = 0.966$ and $r_t = 0.965$) and ($r_a = 0.951$ and $r_a = 0.951$) values obtained by SVM2 and ANN2 techniques show that measures and estimates are well correlated. Inclusion of temperatures (Tmax and Tmin) increased values of correlation coefficients for SVM2 and ANN2 combinations in relation to SVM1 and ANN1 combinations. The combination of H_O, n/N, Tmax and Tmin variables in SVM2 and ANN2 models represent the joining of all variables of A-P, Hargreaves-Samani and Bristow-Campbell equations. The latter two equations use Tmin and Tmax variables (temperature range) to estimate H (Hargreaves and Samani, 1982; Bristow and Campbell, 1984). ANN2 showed the highest r amplitude variation between models of Table 4, r_t value of 0.924 increased to 0.965, while $r_a=0.932$ increased to 0.951. SVM2 performed better than the ANN2 network in both validation conditions.

For SVM3 and ANN3 combinations (Fig. 5e, f), adding precipitation improved the performance with a small gain in r values: $r_t = 0.969$ and $r_t = 0.955$, and $r_a = 0.964$ and $r_a = 0.952$, respectively. Similarly, for SVM4 and ANN4 models (Fig. 5g, h), adding relative humidity (%) in ANN3 and SVM3 combination kept $r_t = 0.970$ and $r_t = 0.958$ and $r_a = 0.963$ and $r_a = 0.951$ values, respectively, very close to those obtained with SVM2 and ANN2 in both validation conditions.

* SVM   o ANN   —— Regression$_{SVM}$ —— Regression$_{ANN}$
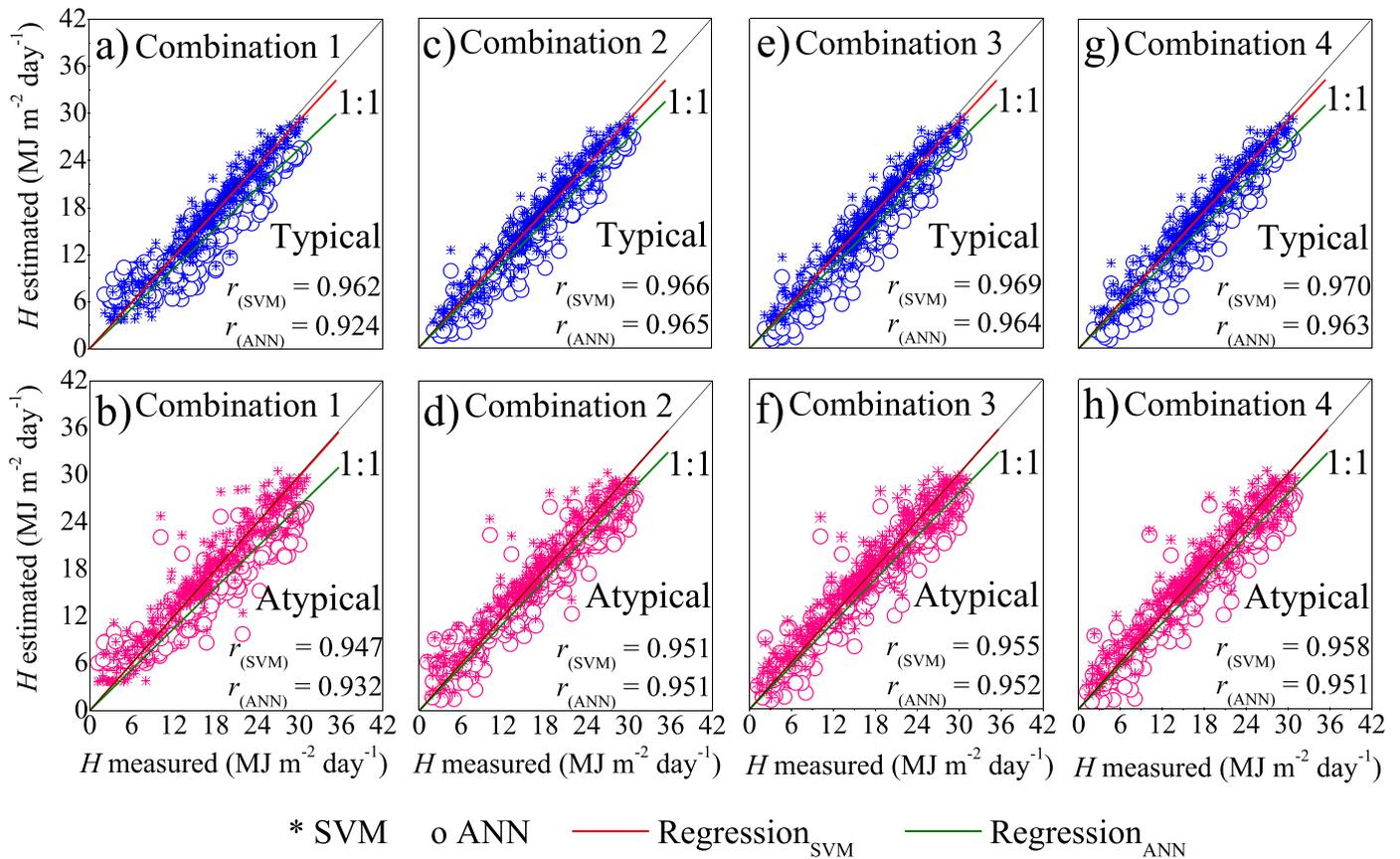
Fig. 5. (a-h). Correlations between values estimated by ML models and measured *H* values.

SVM   ANN
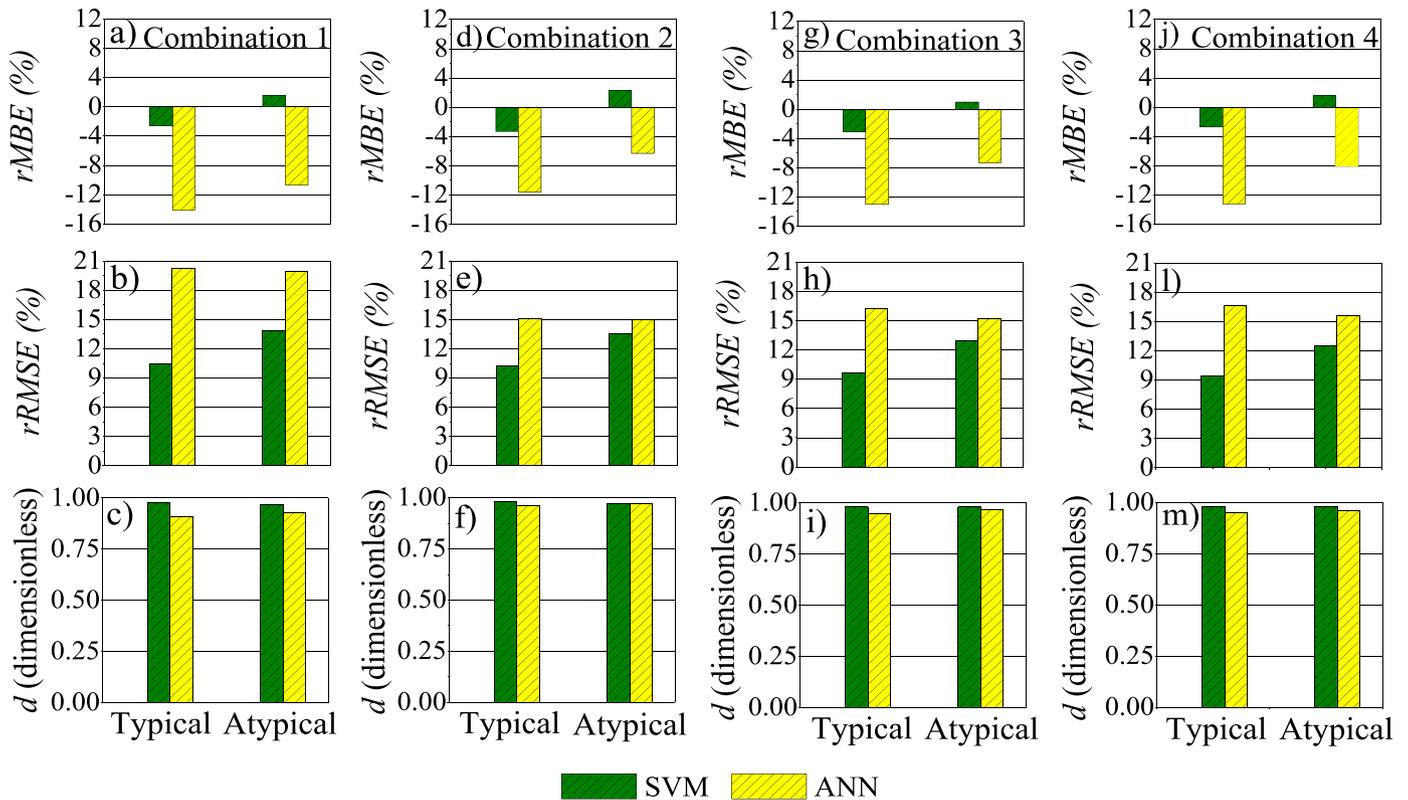
Fig. 6. (a-m). Statistical indicators *rMBE* (%), *rRMSE* (%) and *d* Willmott in the validation.

Table 5
Statistical indicators obtained for SVM4 and ANN4 in Botucatu for typical and atypical years and other locations.

Author	Locations(N°)	Latitude	Models	rMBE(%)	MBE(MJ m⁻² day⁻¹)	rRMSE(%)	RMSE(MJ m⁻² day⁻¹)	r	r²
Chen et al. (2013)	3 (China)	38°54′ – 41°44′ N	SVM	–	–	13.40–17.40	1.79–2.38	–	–
Ramedani et al. (2014)	Teerã (Iran)	35°0′ N	SVM	–	–	–	3.3	–	0.89
Chen et al. (2015)	32 (China)	22°37′ – 33°01′48″ N	SVM	–	–	–	0.84–2.66	–	–
Piri et al. (2015)	2 (Iran)	29°30′ – 37°28′ N	SVM	–	–	–	1.63–2.07	–	0.72–0.93
Mohammadi et al. (2015a)	Isfahan (Iran)	32°39′41″ N	SVM	–	–	9.0	2.01	–	0.91
Mohammadi et al. (2015b)	Bandar Abbas (Iran)	27°13′ N	SVM	–	–	3.69–10.36	0.66–1.85	–	0.80–0.97
Urraca et al. (2015)	4 (Spain)	36°54′ – 37°53′24″ N	SVM	–	–	–	2.39	–	0.92
Present study	*Botucatu (Brazil)*	*22°53′ S*	*SVM4*	*(−2.70)−1.60*	*(−0.47)−0.28*	*9.40–12.50*	*1.64–2.14*	*0.96–0.97*	*0.92–0.94*
Elizondo et al. (1994)	3 (United States)	29°39′ – 35°39′ N	ANN	–	–	–	2.92–3.64	–	0.52–0.74
Bocco et al. (2006)	Cordoba (Argentina)	31°26′ S	ANN	–	–	–	3.14–3.88	0.86–0.92	0.96–0.99
Bosch et al. (2008)	Granada (Spain)	37°10′35″ N	ANN	(−1.20)−2.10	–	5.0–7.50	–	–	0.98
Lam et al. (2008b)	40 (China)	23°23′ – 49°13′ N	ANN	(−16.90)−18.60	(−3.19)−2.19	9.10–20.50	1.40–4.01	–	0.81–0.96
Wan et al. (2008)	41 (China)	22°18′ – 49°13′ N	ANN	(−16.0)−18.90	(−3.08)−2.14	8.20–24.30	1.27–4.00	–	–
Fortin et al. (2008)	Montreal (Canada)	45°0′ – 47°18′ N	ANN	–	–	–	3.83–5.45	–	–
Bocco et al. (2010)	Salta (Argentina)	24°54′ S	ANN	–	–	–	1.66–2.97	–	0.73–0.92
Rahimikhoob (2010)	Ahwaz (Iran)	31°20′ N	ANN	–	–	–	2.53	–	0.89
Linares-Rodríguez et al. (2011)	Andalusia (Spain)	37°32′39″ N	ANN	–	−0.47	16.4	2.88	0.94	–
Landeras et al. (2012)	Basque Country (Spain)	42°38′24″ – 42°51′36″ N	ANN	–	–	–	2.93	–	–
Mejdoul and Belouaggadia (2013)	8 (Morocco)	30°19′48″ – 35°09′ N	ANN	–	–	–	1.23–1.50	0.98	–
Olatomiwa et al. (2015)	3 (Nigeria)	7°57′36″ – 11°49′48″ N	ANN	–	–	–	0.55–2.98	0.37–0.86	0.13–0.74
Mohammadi et al. (2015b)	Bandar Abbas (Iran)	27°13′ N	ANN	–	–	10.08–11.44	1.81–2.05	–	0.75–0.81
Lyra et al. (2015)	Alagoas (Brazil)	9°1′48″ – 10°12′36″ S	ANN	–	–	10.10	1.97	–	0.75
Gupta and Singhal (2016)	4 (India)	21°08′48″ – 26°55′20″ N	ANN	–	–	–	0.24–0.73	–	0.86–0.93
Urraca et al. (2015)	4 (Spain)	36°54′ – 37°53′24″ N	ANN	–	–	–	2.38	–	0.92
Present study	*Botucatu (Brazil)*	*22°53′ S*	*ANN4*	*(−13.20)−(−8.10)*	*(−2.30)−(−1.38)*	*15.60–16.60*	*2.68–2.89*	*0.95–0.96*	*0.90–0.92*

N° = Number of locations.
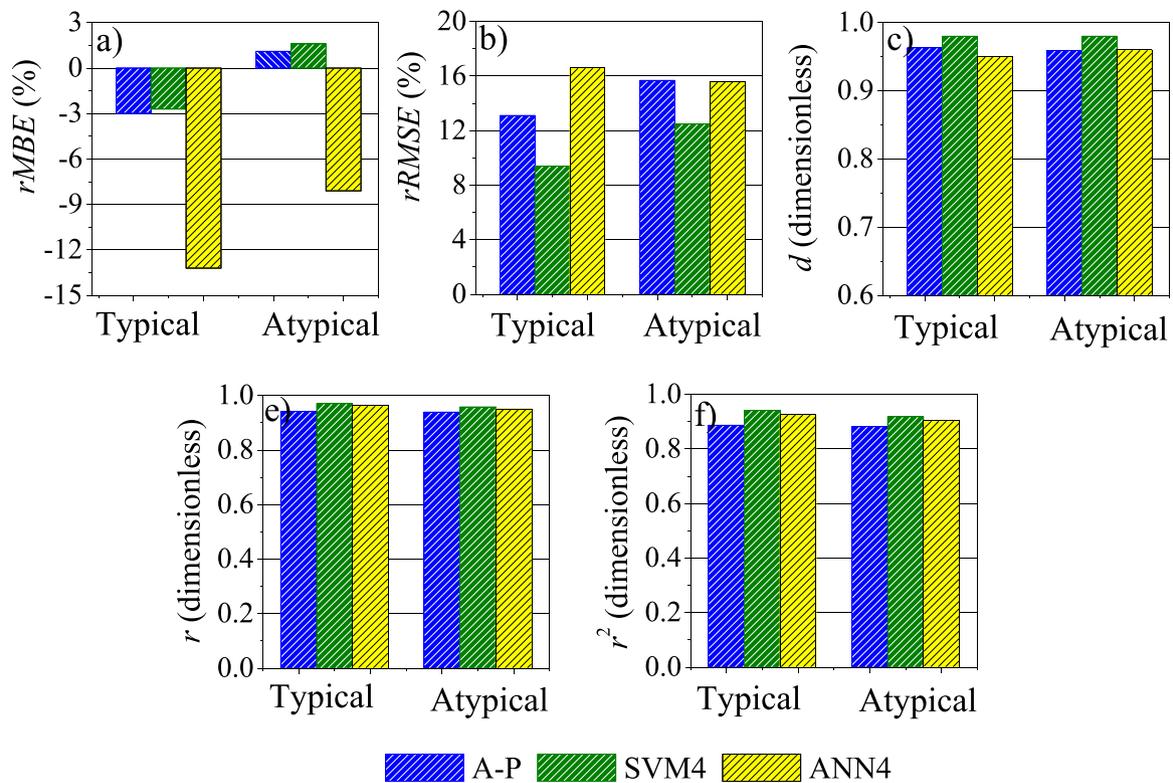
Fig. 7. (a-f). Statistical indicators *rMBE, rRMSE, d* Willmott, *r* and *r²* for typical and atypical years.

Values of *r* obtained for the four models (Fig. 5) in both conditions of validation show that as of the second combination, the other ones can estimate *H* with equal or similar accuracy. Therefore, for comparisons between the Angstrom-Prescott model and SVM and ANN techniques, SVM4 and ANN4 combinations were chosen.

Statistical indicators *MBE, RMSE* and *d* Willmott obtained from the comparison of estimated and measured *H* values, respectively for typical and atypical years, are shown in Fig. 6(a-m).

For combination 1, *rMBE* values (Fig. 6a) of validation with typical and atypical years, respectively, show that $rMBE^t_{ANN1} = -14.1\%$ and $rMBE^a_{ANN1} = -10.7\%$. The *rRMSE* values (Fig. 6b) were $rRMSE^t_{SVM1} = 10.5\%$ and $rRMSE^a_{SVM1} = 13.9\%$, while $rRMSE^t_{ANN1} = 20.3\%$ and $rRMSE^a_{SVM1} = 19.9\%$, indicating that on average, $rRMSE_{SVM1}$ values are 7.9% lower than $rRMSE_{ANN1}$ values. The *d* Willmott concordance index (Fig. 6c) for *SVM1* was superior in the sequence for $d^t = 0.98$ and $d^a = 0.97$ (average of 0.975 ± 0.007), whereas for $ANN1 d^t = 0.91$ and $d^a = 0.93$ (average of 0.92 ± 0.014), respectively.

For combination 2 (Fig. 6d), *rMBE* values show $rMBE^t_{SVM2} = -3.3\%$ and $rMBE^a_{SVM2} = 2.3\%$, while $rMBE^t_{ANN2} = -11.6\%$ and $rMBE^a_{ANN2} = -6.3\%$. The *rRMSE* values and (Fig. 6e) were equal to: $rRMSE^t_{SVM2} = 10.2\%$ and $rRMSE^a_{SVM2} = 13.5\%$ (average of 11.85 ± 2.33), while $rRMSE^a_{ANN2} = 15.1\%$ and $rMSE^a_{ANN2} = 15.0\%$ (average of 15.05 ± 0.071). The *d* Willmott concordance index (Fig. 6f) with SVM2 was superior in sequence for $d^t=0.98$ and $d^a = 0.97$, while for $ANN2 d^t = 0.96$ and $d^a = 0.97$, respectively.

For combination 3, *rMBE* values (Fig. 6g) show $rMBE^t_{SVM3} = -3.1\%$ and $rMBE^a_{SVM3} = 2.2\%$, while for $rMBE^a_{ANN3} = -13.0\%$ and $rMBE^a_{ANN3} = -7.0\%$. The $rMBE_{SVM3}$ values are close to zero. The *rRMSE* values (Fig. 6h) were: $rRMSE^t_{SVM3} = 9.7\%$ and $rRMSE^a_{SVM3} = 13.0\%$ while for $rRMSE^t_{ANN3} = 16.3\%$ and $rRMSE^a_{ANN3} = 15.6\%$. On average, $rRMSE_{SVM3}$ value was 4.6% lower than $rRMSE_{ANN3}$. The *d* Willmott concordance index (Fig. 6i) with SVM3 was equal for $d^t = 0.98$ and $d^a = 0.98$, while for ANN1 $d^t = 0.95$ and $d^a = 0.97$, respectively.

For combination 4, *rMBE* values (Fig. 6j) show SVM4: $rMBE^t_{SVM4} = -2.7\%$ and $rMBE^a_{SVM4} = 1.6\%$, while for *ANN4*: $rMBE^t_{ANN4} = -13.2\%$ and $rMBE^a_{ANN4} = -8.1\%$. The *rRMSE* values (Fig. 6l) were:

$rRMSE^t_{SVM4} = 9.5\%$ and $RMSE^a_{SVM4} = 12.5\%$, while $rRMSE^t_{ANN4} = 16.6\%$ and $rRMSE^a_{ANN4} = 15.6\%$ (average of 16.1 ± 0.707). The *d* Willmott concordance index (Fig. 5m) with SVM4 was equal to $d^t = 0.98$ and $d^a = 0.98$ (average of 0.98 ± 0), while for ANN4 $d^t = 0.95$ and $d^a = 0.96$ (average of 0.955 ± 0.007), respectively.

The comparison of *r, rMBE, rRMSE* and *d* statistical indicators of the training validation of ML techniques with that of A-P^t model shows that the SVM1 combination has approximately the same performance as that of the A-P^t model, and higher than that of ANN1 in the two validation conditions. On the other hand, the A-P^t has better performance than that of ANN1 in the two validation conditions. In practical use, the choice between the A-P method and ANN1 combination should be for the statistical model, whose indexes are more favorable for *H* estimates. Concerning the A-P model and the SVM1 combination, the choice is at the discretion of the user. The statistical method is the most preferable, as it is more practical. However, computational software techniques have been used because of their good performance and continuous upgrade.

The *rMBE* values obtained with SVM (below 4.0%) are more statistically significant than those of ANN (below 14.0%), with better results for models 2, 3 and 4 (ANN and SVM). Similarly, *rRMSE* values obtained with the SVM technique (below 15.0%) were also statistically more significant than those with ANN (below 20.0%), with better results for models 2, 3 and 4. The *d* Willmott values obtained with SVM are closer to 1 than ANN values, with better results for models 2, 3 and 4. In general, statistical indicators *rMBE, rRMSE* and *d* Willmott obtained for the four models (Fig. 6) show that from SVM2 and ANN2, models can also estimate *H* values with precision and accuracy, thus justifying the choice of SVM4 and ANN4 models to compare with the Angstrom-Prescott model.

3.4. Performance of models (SVM and ANN) and comparison with literature

Table 5 shows a comparison of the statistical indicators *MBE* (MJ m^{-2} day^{-1}), *rMBE* (%), *RMSE* (MJ m^{-2} day^{-1}), *rRMSE* (%) *r* or *r²*

Table 6
Statistical indicators for ML (SVM and ANN) and A-P models in Botucatu and other locations.

Authors	Locations(N°)	Latitude	Models	rMBE(%)	MBE(MJ m⁻² day⁻¹)	rRMSE(%)	RMSE(MJ m⁻² day⁻¹)	r	r²
Benghanem and Mellit (2010)	Al-Madinah (Saudi Arabia)	24°55' N	A-P	–	–	–	0.0002	0.97	–
Benghanem and Mellit (2010)	//	//	SVM	–	–	–	0	0.98	–
Chen et al. (2013)	3 (China)	38°54' – 41°44' N	A-P	–	–	14.91–19.42	1.99–2.65	–	–
Chen et al. (2013)	//	//	SVM	–	–	13.37–17.43	1.79–2.38	–	–
Mohammadi et al. (2015a)	Isfahan (Iran)	32°39'41'' N	A-P	–	–	13.86	2.67	–	0.85
Mohammadi et al. (2015a)	//	//	SVM	–	–	9,0	2.01	–	0.91
Piri et al. (2015)	2 (Iran)	29°30' – 37°28' N	A-P	–	–	–	3.38–2.81	–	0.78–0.79
Piri et al. (2015)	//	//	SVM	–	–	–	1.63–2.07	–	0.72–0.93
Wan et al. (2008)	41 (China)	22°18' – 49°13' N	A-P	(–16.2)–18.8	(–3.13)–2.13	6.8–31.3	1.23–3.78	–	–
Wan et al. (2008)	//	//	ANN	(–16.1)–19.7	(–3.10)–2.14	8.2–30.3	1.24–4.00	–	–
Present study	**Botucatu (Brazil)**	**22°53' S**	**A-P**	**(–2.98)–1.10**	**(–0.52)– 0.19**	**13.11–15.68**	**2.28–2.69**	**0.94–0.94**	**0.88–0.88**
Present study	//	//	**SVM**	**(–2.7)–1.6**	**(–0.47)– 0.28**	**9.4–12.5**	**1.64–2.14**	**0.96–0.97**	**0.92–0.94**
Present study	//	//	**ANN**	**(–13.2)–(–8.1)**	**(–2.30)– 1.38**	**15.6–16.6**	**2.68–2.89**	**0.95–0.96**	**0.90–0.92**

N° = Number of locations.

obtained in both validation conditions, typical and atypical years, between measures and estimates with SVM4 and ANN4 in Botucatu, and in various locations in the world. The studies listed in Table 5 estimate H based on ANN and SVM using different meteorological variables as input parameters.

Comparison of indexes for typical and atypical years shows that SVM4 presents more statistically significant results than ANN4, i.e., it has better performance estimating H. $rMBE^t_{SVM4} = -2.7\%$ and $rMBE^a_{SVM4} = 1.6\%$ values are smaller than $rMBE^t_{ANN4} = (-13.2\%)$ and $rMBE^a_{ANN4} = (-8.16\%)$; values $rRMSE^t_{SVM4} = 9.4\%$ and $rRMSE^a = 12.5\%$ are smaller than $rRMSE^t_{ANN4} = 15.6\%$ and $rRMSE^a_{ANN4} = 16.6\%$; the d Willmott values $d^t_{SVM4} = 0.98$ and $d^a_{SVM4} = 0.98$ are greater than those of $d^t_{ANN4} = 0.95$ and $d^a_{ANN4} = 0.96$; and the $r^t_{SVM4} = 0.970$ and $r^a_{SVM4} = 0.918$ values are greater than $r^t_{ANN4} = 0.95$ and $r^a_{ANN4} = 0.96$ values and $(r^2)^t_{SVM4} = 0.941$ and $(r^2)^a_{SVM4} = 0.842$ values are greater than $(r^2)^a_{ANN4} = 0.90$ and $(r^2)^a_{ANN4} = 0.92$.

MBE, *RMSE*, r and r^2 values obtained in this study (Table 5) with the SVM4 and ANN4 techniques are inferior or at the same order of magnitude as, or superior to the results of SVM and ANN obtained in other locations in the North and South hemispheres. Many factors, such as technique training, types of ML programming and validation process of techniques influenced the variability of the statistical indicators (mainly *rRMSE*, common to all studies in Table 5) obtained with the SVM4 and ANN4 techniques and SVM and ANN techniques for other locations. In the training of ML techniques, the input of different variables (meteorological and astronomical) provide different estimates by the models. Also, the time duration of the data base used in the technique training is important in the modeling process, because a more extended data base provides more details about the influence of the local climate on measures, and improves the ML model performance. The types of ML models cause variation in the estimates because the computational programs can operate with different algorithms and mathematical functions. Many studies (Table 5) used different algorithms and mathematical functions in the SVM and ANN techniques (Landeras et al., 2012; Chen et al., 2013; Lyra et al., 2015; Mohammadi et al., 2015b; Urraca et al., 2015). Similar to the process of technique training, the technique validation is also a factor which affects *RMSE* variability. The validation depends on the criterion and time duration of the data base of the measures to be compared with the estimates of SVM and ANN techniques. In this study, the criterion used for the validation by two bases, typical and atypical years, differs from the methodologies adopted in the studies of Table 5.

3.5. Comparison of the performance of the A-P Model and ML (SVM and ANN) in Botucatu and other locations

Fig. 7(a-f) shows the comparison of statistical indicators *rMBE*, *rRMSE*, d Willmott, r and r^2 of the A-P, ANN and SVM models obtained in Botucatu.

The comparison among validation indicators shows that the statistical performance of SVM4 is superior to that of the A-P model. $MBE^t_{SVM4} = -2.7\%$ and $MBE^a_{SVM4} = 1.6\%$ values are at the same order of magnitude as $MBE^t_{A-P} = -3.0\%$ and $MBE^a_{A-P} = 1.1\%$ values (Fig. 7a); the $RMSE^t_{SVM4} = 9.4\%$ and $RMSE^a_{SVM4} = 12.5\%$ values are smaller than $RMSE^t_{A-P} = 13.1\%$ and $RMSE^a_{A-P}$ values in both validation bases (Fig. 7b); d Willmott values $d^t_{SVM4} = 0.98$ and $d^a_{SVM4} = 0.98$ are greater than $d^t_{A-P} = 0.95$ and $d^a_{A-P} = 0.96$ values (Fig. 7c); $r^t_{SVM4} = 0.970$ and $r^a_{SVM4} = 0.958$ values are greater than $r^t_{A-P} = 0.942$ and $r^a_{A-P} = 0.939$ values, respectively (Fig. 7d) and $(r^2)^t_{SVM4} = 0.941$ and $(r^2)^a_{SVM4} = 0.918$ values are greater than $(r^2)^t_{A-P} = 0.887$ and $(r^2)^a_{A-P} = 0.882$ values (Fig. 7e).

Statistical indicators obtained in the SVM4 validation, superior to those of the A-P model, show that meteorological input variables in the machine learning technique are likely to improve the performance of H estimates, which is the function of the artificial intelligence model. Several recent studies in the literature have shown that when the A-P
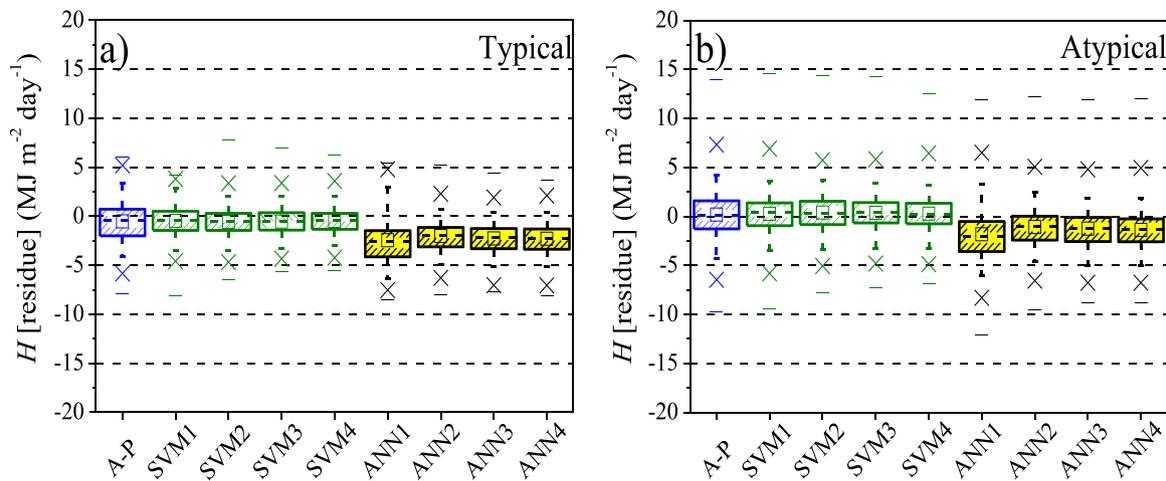
Fig. 8. (a, b). Box plots of deviations of A-P, SVM and ANN models. a) Typical year and b) Atypical year.

statistical model is modified through the introduction of new meteorological variables (multiple regressions), it may have better performance than that of the SVM technique obtained in this study (Amororox, 2011b; Adeala et al., 2015; Kutty et al., 2015; Coulibaly and Quedraogo, 2016).

The results of statistical indicators $rMBE$, $rRMSE$, d Willmott, r and r^2 in the comparison between SVM and A-P models obtained in this study are similar to those by Benghanem and Mellit (2010) in Al-Madinah (Saudi Arabia); Chen et al. (2013) in China; Mohammadi et al. (2015a) and Piri et al. (2015), both in Iran (Table 6). In all studies (Table 6), SVM showed better performance than that of the A-P model.

$rRMSE$ value in Botucatu was lower than that by Chen et al. (2013) in China, and it was at the same order of magnitude as those by Mohammadi et al. (2015a) and Piri et al. (2015) in Iran. It was superior to that by Benghanem and Mellit (2010) in Al-Madinah (Saudi Arabia). The r value in Botucatu was lower than that by Benghanem and Mellit (2010) in Al-Madinah (Saudi Arabia). The results for the indicators obtained in this study are in accordance with and at the same order of magnitude as those by Wan et al. (2008) in 41 cities in China.

Box graphs show residues (H estimated − H measured, MJ m^{-2} day^{-1}) obtained for A-P, SVM and ANN models, in the validation bases of typical (Fig. 8a) and atypical years (Fig. 8b). In each case, the central points are average values of all values and the center line is the median. By default, the box is determined by percentiles 25 and 75.

Narrower boxes show the models with better performance. Accordingly, the sequencing of SVM combinations (SVM1, SVM2, SVM3 and SVM4) corroborates the best performance of SVM4 among the models A-P, SVM and ANN. The performance of the models observed using the Box Plot technique is similar to that of the models A-P, SVM and ANN when analyzed using *MBE* and *RMSE* statistical indexes.

4. Conclusions

The results presented and discussed in this study conclude that results of the A-P model in Botucatu/SP, Brazil, with r^2=0.806 are similar to those found in other locations in the country. Values of the statistical indicators in conditions of typical and atypical validation: $rMBE$ lower than 3.0%; $rRMSE$ lower than 15.68%; d Willmott higher than 0.95 reveal that the A-P can be used to estimate H with precision and accuracy. It would be of great importance that further studies in Botucatu analyze the performance of the model A-P by testing new meteorological variables measured in the earth and atmospheric surface as input variables in multiple regression models.

The ML models (SVM and ANN) evaluated in this study corroborated the results already observed concerning the efficiency of both

techniques to estimate H: the combination 1 (SVM1 and ANN1) and the model A-P, which used the same input variables H_O and (n/N), showed through values of statistical indicators r , $rMBE$, $rRMSE$, d Willmott that the SVM1 technique estimates H at the same order of magnitude as the A-P model, whereas the A-P model estimates H with better accuracy and precision than the ANN1 technique.

The statistical indicators obtained from the combinations 2–4 also confirmed that the input of new variables in the training of techniques improve the performance of the ML model: the combination 2 (SVM2 and ANN2) in relation to combination 1 had better performance by adding the temperatures (T, Tmax and Tmin). Similarly, for H estimation, the combination 3 (SVM3 and ANN3) had improved performance by adding precipitation; the combination 4 (SVM4 and ANN4) had improved performance by adding the relative humidity.

The comparison of values of the statistical indicators $rMBE$, $rRMSE$, d Willmott, r and r^2 obtained in the validation of the A-P models and the combination 4 (SVM4 and ANN4) showed that the SVM4 technique has better performance than that of the A-P model in estimating H; the SVM4 combination has better performance than that of the ANN4 in estimating H; the A-P model has better performance than that of the combination ANN4 in estimating H.

The conclusion about the comparative performance between A-P models and the techniques of machine learning SVM and ANN is valid only for the study location (Botucatu/SP/Brazil). A broader conclusion would require further studies to be held in different locations with several climate conditions and new input variables (astronomical, geographic variables) and atmospheric constituents (clouds, aerosols, water vapor, ozone and other gases).

References

Adeala, A.A., Zhongjie, Z., Enweremadu, C.C., 2015. Evaluation of global solar radiation using multiple weather parameters as predictors for South Africa provinces. Therm. Sci. 19, S495–S509.

Allen, R.G., Pereira, L.S., Raes, D., et al. 1998. Crop Evapotranspiration – Guidelines For Computing Crop Water Requirements. Food and Agriculture Organization of the United Nations (FAO) – Irrigation and drainage paper 56, Rome. p. 50.

Almorox, J., Benito, M., Hontoria, C., 2005. Estimation of monthly Angstrom-Prescott equation coefficients from measured daily data in Toledo, Spain. Renew. Energy 30, 931–936.

Almorox, J., Hontoria, C., Benito, M., 2011a. Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain). Appl. Energy 88, 1703–1709.

Almorox, J., 2011b. Estimating global solar radiation from common meteorological data

in Aranjuez, Spain. Turk. J. Phys. 35, 53–64.

Andrade Júnior, A.S., Noleto, H.D., Silva, M.E., et al., 2012. Coeficientes da equação de Angstrom-Prescott para Parnaíba, Piauí. Comun. Sci. 1 (3), 50–54, (In Portuguese).

Angström, A., 1924. Solar and terrestrial radiation. Q. J. R. Meteorol. Soc. 50, 121–125.

Araújo, C.E., Santos, J.M., Sansigolo, C.A., 2001. Radiação solar global nas cidades de São Paulo, Rio de Janeiro e Brasília no período de 1978–1987. In: Proceedings of the XII Congresso Brasileiro de Agrometeorologia-Fortaleza. [In Portuguese].

Azevedo, P.V., Varejão-Silva, M.A., Vargas, G.A.O., 1981. Zoneamento do Potencial de Energia Solar no Nordeste. UFPB, Campina Grande. [In Portuguese].

Back, A.J., 2005. Determinação dos coeficientes da equação de Angstrom-Prescott para a estimativa da radiação solar global para Urussanga. Sc. Rev. Bras. Agrometeorol. 13 (3), 430–435, (In Portuguese).

Bakirci, K., 2009. Models of radiation with hours of bright sunshine: a review. Renew. Sustain. Energy Rev. 13, 2580–2588.

Bechini, L., Ducco, G., Donatelli, M., et al., 2000. Modeling, interpolation and stochastic simulation in space and time of global solar radiation. Agric. Ecosyst. Environ. 81, 29–42.

Behrang, M.A., Assareh, E., Ghanbarzadeh, A., Noghrehabadi, A.R., 2010. The potential of different Artificial Neural Network (ANN) techniques in daily global solar radiation modeling based on meteorological data. Sol. Energy 84, 1468–1480.

Belúcio, L.P., Silva, A.P.N., Souza, L.R., et al., 2014. Radiação solar global estimada a partir da insolação para Macapá (AP). Revista Brasileira de Meteorologia. 29 (4), 494–504, (In Portuguese).

Benghanem, M., Mellit, A., 2010. Radial Basis Function Network-based prediction of global solar radiation data: application for sizing of a stand-alone photovoltaic system at Al-Madinah, Saudi Arabia. Energy 35, 3751–3762.

Berusky, G.C., Pereira, A.B., Sentelhas, P.C., 2015. Desempenho de diferentes modelos de estimativa da radiação solar global em Ponta Grossa. Pr. Rev. Bras. Meteorol. 30 (2), 83–91, (In Portuguese).

Blanco, F.F., Sentelhas, P.C., 2002. Coeficientes da equação de Angstrom-Prescott para estimativa da insolação para Piracicaba, SP. Rev. Bras. Agrometeorol. 10 (2), 295–300, (In Portuguese).

Bocco, M., Ovando, G., Sayago, S., 2006. Development and evaluation of neural network models to estimate daily solar radiation at Córdoba, Argentina. Pesqui. Agropecu. Bras. 41, 179–184.

Bocco, M., Willington, E., Arias, M., 2010. Comparison of regression and neural networks. Chil. J. Agric. Res. 70 (3), 428–435.

Bojanowski, J.S., Donatelli, M., Skidmore, A.K., et al., 2013. An auto-calibration procedure for empirical solar radiation models. Environ. Model. Softw. 49, 118–128.

Bosch, J.L., López, G., Batlles, F.J., 2008. Daily solar irradiation estimation over a mountainous area using artificial neural networks. Renew. Energy 33, 1622–1628.

Bristow, K.L., Campbell, G.S., 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agric. For. Meteorol. 31 (2), 159–166.

Carvalho, D.F., Silva, D.G., Souza, A.P., et al., 2011. Coeficientes da equação de Angstrom-Prescott e sua influência na evapotranspiração de referência em Seropédica. Rj. Rev. Bras. Eng. Agríc. Ambient. 15 (8), 838–844, (In Portuguese).

Castillo, C.P., Silva, F.B., Lavalle, C., 2016. An assessment of the regional potential for solar power generation in EU-28. Energy Policy 88, 86–99.

Chaves, M.A., Escobedo, J.F., 2000. A software to process daily solar radiation data. Renew. Energy 19, 339–344.

Chen, R., Ersi, K., Yang, J., et al., 2004. Validation of five global radiation models with measured daily data in China. Energy Convers. Manag. 45, 1759–1768.

Chen, J.-L., Li, G.-S., Wu, S.-J., 2013. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. Energy Convers. Manag. 75, 311–318.

Chen, J.-L., Li, G.-S., Xiao, B.B., 2015. Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. Energy Convers. Manag. 89, 318–329.

Coulibaly, O., Quedraogo, A., 2016. Correlation of Global Solar Radiation of eight synoptic stations in Burkina Faso based on linear and multiple linear Regression methods. J. Sol. Energy. http://dx.doi.org/10.1155/2016/7870907.

Dai, Q., Fang, X., 2014. A simple model to predict solar radiation under clear sky conditions. Adv. Space Res. 53, 1239–1245.

Dallacort, R., Freitas, P.S.L., Gonçalves, A.C.A., et al., 2004. Determinação dos coeficientes da equação de Angstrom para a região de Palotina, Estado do Paraná. Acta Sci. Agron. 26 (3), 329–336, (In Portuguese).

Dantas, A.A.A., Carvalho, L.G., Ferreira, E., 2003. Estimativa da radiação solar global para a região de Lavras, MG. Ciência Agrotecnol. 27, 1260–1263, (In Portuguese).

Dornelas, K.D.S., Silva, C.L., Oliveira, C.A.S., 2006. Coeficientes médios da equação de Angstrom-Prescott, radiação solar e evapotranspiração de referência em Brasília. Pesqui. Agropecu. Bras. 41 (8), 1213–1219, (In Portuguese).

Elizondo, D., Hoogenboom, G., Mcclendon, R.W., 1994. Development of a neural network model to predict daily solar radiation. Agric. For. Meteorol. 71, 115–132.

El-Sebaii, A.A., Al-Ghamdi, A.A., F. S. Al-Hazmi, F.S., et al., 2009. Estimation of global solar radiation on horizontal surfaces in Jeddah, Saudi Arabia. Energy Policy 37, 3645–3649.

Escobedo, J.F., Teramoto, E.T., Oliveira, A.P., et al., 2012. Equações de estimativa das frações solar direta (k$_d$^{h}) e difusa (k$_d$) em função do índice de claridade (kt) e razão de insolação (n/N). Av. En. Energ. Renov. Medio Ambient. 16, (In Portuguese).

Fortin, J., Anctil, F., Parent, L., et al., 2008. Comparision of empirical daily surface incoming solar radiation models. Agric. For. Meteorol. 148, 1332–1340.

Gueymard, C.A., 2003. Direct solar transmittance and irradiance predictions with broadband models. Part II: validation with high-quality measurements. Sol. Energy 74 (5), 381–395.

Gupta, R., Singhal, S., 2016. Prediction of Global Solar Radiation in India using Artificial

Neural Network. J. Sustain. Dev. Energy Water Environ. Syst. 4, 94–106.

Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. J. Irrig. Drain. Eng. 108, 225–230.

Haykin, S., 1998. Neural Networks: A Comprehensive Foundation 2nd ed.. Prentice Hall, Hamilton, 897.

Hsiao, T., Heng, L., Steduto, P., et al., 2008. Aqua crop – the FAO crop model to simulate yield response to water: iii. Parameterization and testing for maize. Agron. J. 101, 448–459.

Hunt, L.A., Kuchar, L., Swanton, C.J., 1998. Estimation of solar radiation for use crop modeling. Agric. For. Meteorol. 91, 293–300.

Iziomon, M.G., Mayer, H., 2001. Performance of solar radiation models – a case study. Agric. For. Meteorol. 110, 1–11.

Jamieson, P.D., Porter, J.R., Wilson, D.R., 1991. A test of the computer simulation model ARC – WHEAT1 on wheat crops grown in New Zealand. Field Crops Res. 27, 337–350.

Jiang, Y., 2008. Prediction of monthly mean daily diffuse solar radiation using artificial neural network's and comparison with other empirical models. Energy Policy 36, 3833–3837.

Kaushika, N.D., Tomar, R.K., Kaushik, S.C., 2014. Artificial Neural Network model based on interrelationship of direct, diffuse and global solar radiations. Sol. Energy 103, 327–342.

Kutty, H.A., Masral, M.H., Rajendran, P., 2015. Regression Model to Predict Global Solar Irradiance in Malaysia. Int. J. Photoenergy. http://dx.doi.org/10.1155/2015/347023.

Lam, J.C., Wan, K.K.W., Lau, C.C.S., et al., 2008a. Climatic influences on solar modeling in China. Renew. Energy 33, 1591–1604.

Lam, J.C., Wan, K.K.W., Yang, L., 2008b. Solar radiation modeling using ANNs for different climates in China. Energy Convers. Manag. 49, 1080–1090.

Landeras, G., López, J.J., Kisi, O., Shiri, J., 2012. Comparison of Gene Expression Programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain). Energy Convers. Manag. 62, 1–13.

Li, H., Ma, W., Lian, Y., et al., 2011. Global solar radiation estimation with sunshine duration in Tibet, China. Renew. Energy 36, 3141–3145.

Li, M., Fan, L., Liu, H., et al., 2012. Impact of time interval on the Angstrom-Prescott coefficients and their interchangeability in estimating radiation. Renew. Energy 44, 431–438.

Li, M., Tang, X., Wu, W., et al., 2013. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. Energy Convers. Manag. 70, 139–148.

Lima, F.J.L., Martins, F.R., Pereira, E.B., et al., 2016. Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks. Renew. Energy 87, 807–818.

Linares-Rodríguez, A., Ruiz-Arias, J.A., Pozo-Vásquez, D., et al., 2011. Generation of synthetical daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. Energy 36, 5356–5365.

Liu, D.L., Scott, B.J., 2001. Estimation of solar radiation in Australia from rainfall and temperature observations. Agric. For. Meteorol. 106, 41–59.

Liu, X., Mei, X., Li, Y., et al., 2009. Calibration of the Angstrom-Prescott coefficients (a, b) under different times scales and their impacts in estimating global solar radiation in the Yellow River Basin. Agric. For. Meteorol. 149, 697–710.

Lyra, G.B., Zanetti, S.S., Santos, A.A.R., et al., 2015. Estimation of monthly global solar irradiation using the Hargreaves-Samani model and na artificial neural network for the state of Alagoas in northeastern Brazil. Theor. Appl. Climatol.. http://dx.doi.org/10.1007/s00704-015-1541-8.

Madkour, M.A., El-Metwally, M., Hamed, A.B., 2006. Comparative study on different models for estimation of direct normal irradiance (DNI) over Egypt atmosphere. Renew. Energy 31, 361–382.

Manzano, A., Martín, M.L., Valero, F., et al., 2015. A single method to estimate the daily global solar radiation from monthly data. Atmos. Res. 166, 70–82.

Martim, C.C., Souza, A.P., Paulino, J., et al., 2014. Coeficientes de Angstrom-Prescott para a Região Norte do Mato Grosso: primeira aproximação. In: Proceedings of the XLII Congresso Brasileiro de Engenharia Agrícola-Campo Grande. [In Portuguese].

Martins, F.R., Pereira, E.B., Silva, S.A.B., et al., 2008. Solar energy scenarios in Brazil, Part one: resource assessment. Energy Policy 36, 2853–2864.

Martins, F.R., Abreu, S.L., Pereira, E.B., 2012. Scenarios for solar thermal energy applications in Brazil. Energy Policy 48, 640–649.

Mejdoul, R.M.T., Belouaggadia, N., 2013. Artificial neural network based prediction model of daily global solar radiation in Morocco. J. Renew. Sustain. Energy 5 (063137), 1–9.

Mohammadi, K., Shamshirband, S., Anisi, M.H., et al., 2015a. Support vector regression based prediction of global solar radiation on a horizontal surface. Energy Convers. Manag. 91, 433–441.

Mohammadi, K., Shamshirband, S., Tong, C.W., et al., 2015b. A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers. Manag. 92, 162–171.

Mota, F.S., Beirsdorf, M.I.C., Acosta, M.J.C., 1977. Estimates of solar radiation in Brazil. Agric. Meteorol. 18, 241–254.

Nicácio, R.M., Souza, J.L., Bernardo, S.O., 2001. Estimativa da irradiância solar global para Maceió utilizando o modelo linear de Angstrom-Prescott. In: Proceedings of the XII Congresso Brasileiro de Agrometeorologia – Fortaleza. [In Portuguese].

Olatomiwa, L., Mekhilef, S., Shamshirband, S., et al., 2015. A support vector machine-firefly algorithm-based model for global solar radiation prediction. Sol. Energy 115, 632–644.

Oliveira, A.P., Escobedo, J.F., Machado, A.J., Soares, J., 2002. Correlation model of diffuse solar-radiation applied to the city of São Paulo, Brazil. Appl. Energy 71,

59–73.

Oliveira, A.P., Soares, J., Boznar, M.Z., Mlakar, P., Escobedo, J.F., 2006. An application of Neural Network technique to correct the dome temperature effects on pyrgeometer measurements. J. Atmos. Ocean. Technol. 23, 80–89.

Pacheco, N.A., Bastos, T., 2002. Estimativa da radiação solar global diária em Capitão Poço, PA através da equação de Angstrom. In: Proceedings of the XII Congresso Brasileiro de Meteorologia-Foz do Iguaçu. [In Portuguese].

Park, J., Das, A., Park, J.H., 2015. A new approach to estimate the spatial distribution of solar radiation using topographic factor and sunshine duration in South Korea. Energy Convers. Manag. 101, 30–39.

Pereira, S.T., Santos, W.C., Amorim, J.S., et al., 2010. Estimativa da radiação solar global para a Região de Pedra Azul. Encicl. Biosf. 6 (11), 1–9, (In Portuguese).

Pilau, F.G., Barbieri, V., Marin, F.R., 2007. Coeficientes da equação de Angstrom-Prescott para estimativa da irradiância solar global na região de Araras. Sp. Rev. Bras. Agrometeorol. 15 (1), 109–113, (In Portuguese).

Piri, J., Shamshirband, S., Petkovic, D., et al., 2015. Prediction of the solar radiation on the Earth using support vector regression technique. Infrared Phys. Tecnol. 68, 179–185.

Platt, J., 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Redmond: Microsoft Research, Tech Report. p. 21.

Podestá, G.P., Núñez, L., Villanueva, C.A., et al., 2004. Estimating daily solar radiation in the Argentine Pampas. Agric. For. Meteorol. 123, 41–53.

Prescott, J.A., 1940. Evaporation from a water surface in relation to solar radiation. Trans. R. Soc. Sci. Aust. 64, 114–118.

Quej, V.H., Almorox, J., Arnaldo, J.A., Saito, L., 2017. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J. Atmos. Sol. Terr. Phys.. http://dx.doi.org/10.1016/j.jastp.2017.02.002.

Rahimikhoob, A., 2010. Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. Renew. Energy 35, 2131–2135.

Reda, I.M., Myers, D.R., StoffeL, T.L., 2008. Uncertainty estimate for the outdoor calibration of solar pyranometers: NCSLI measure. J. Meas. Sci. 3 (4), 58–66.

Rehman, S., Mohandes, M., 2008. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. Energy Policy 36, 571–576.

Ramedani, Z., Omid, M., Keyhani, A., et al., 2014. Potential of radial basis function based support vector regression for global solar radiation prediction. Renew. Sustain. Energy Rev. 39, 1005–1011.

Sabzipavar, A.A., Mousavi, R., Marofi, S., et al., 2013. An improved estimation of the Angstrom-Prescott radiation coefficients for the FAO 56 Penman-Monteith evapotranspiration method. Water. Manag. 27, 2839–2854.

Santos, R., Hernandez, F.B.T., Fioravanti, C.D., et al., 2003. Estimativa da radiação solar global diária em Ilha Solteira, São Paulo. In: XXXII Congresso Brasileiro de Engenharia Agrícola – 28 de julho a 01 de agosto. [In Portuguese].

Santos, C.M., Souza, J.L., Ferreira Júnior, R.A., et al., 2014. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. Energy 71, 338–398.

Santos, C.M., Escobedo, J.F., 2016. Temporal variability of atmospheric turbidity and DNI attenuation in the sugarcane region, Botucatu/SP. Atmos. Res. 181, 312–321.

Santos, C.M., Escobedo, J.F., Tadao, E.T., Silva, S.H.M.G., 2016. Assessment of ANN and SVM models for estimating normal direct irradiation (H_b). Energy Convers. Manag. 126, 826–836.

Shamshirband, S., Petkovic, D., Saboohi, H., et al., 2014. Wind turbine power coefficient estimation by soft computing methodologies: comparative study. Energy Convers. Manag. 81, 520–526.

Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., 2000. Improvements to the SMO algorithm for SVM Regression. IEEE Trans. Neural Netw. 11 (5).

Smola, A.J., Schölkopf, B., 1998. A Tutorial on Support Vector Regression. Royal Holloway College, London, U. K., (Neuro COLT Tech. Rep. TR 1998-030).

Smola, A.J., Schölkopf, B., 2004. A tutorial on Support Vector Regression. Stat. Comput. 14, 199–222.

Soares, J., Oliveira, A.P., Boznar, M.Z., Mlakar, P., Escobedo, J.F., Machado, A.J., 2004. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique. Appl. Energy 79, 201–214.

Souza, J.L., Nicácio, R.M., Moura, M.A.L., 2005. Global solar radiation measurements in Maceió, Brazil. Renew. Energy 30, 1203–1220.

Souza, J.L., Lyra, G.B., Santos, C.M., et al., 2016. Empirical models of daily and monthly global solar irradiation using sunshine duration for Alagoas State, Northeastern Brazil. Sustain. Energy Technol. Assess. 14, 35–45.

Texeira, A.H.C., Silva, T.G.F., Reis, V.C. et al., 2002. Radiação solar global e insolação no município de Juazeiro-BA. In: Proceedings of the XII Congresso Brasileiro de Meteorologia – Foz do Iguaçu. [In Portuguese].

Torres, C.J.F., Silva, N.L., Barros, F.M., et al., 2010. Determinação dos coeficientes do modelo de Angstrom-Prescott para a Região de Canavieiras, Estado da Bahia. Encicl. Biosf. 6 (11), 1–7, (In Portuguese).

Trnka, M., Zalud, Z., Eitzinger, J., Dubrovsky, M., 2005. Global solar radiation in Central Europe lowlands estimated by various empirical formulae. Agric. For. Meteorol. 131, 54–76.

Tymvios, F.S., Jacovides, C.P., Michaelides, S.C., et al., 2005. Comparative study of Angstrom and artificial neural networks methodologies in estimating global solar radiation. Sol. Energy 78, 752–762.

Urraca, R., Antonanzas, J., Martínez-de-Pizón, F.J., Antonanzas-Torres, F., 2015. Estimation of solar global irradiation in remote areas. J. Renew. Sustain. Energy 7, 023136. http://dx.doi.org/10.1063/1.4919084.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Vapnik, V., N., 1998. Statistical Learning Theory. Wiley, New York.

Wan, K.K.W., Tang, H.L., Yang, L., et al., 2008. An analysis of thermal and solar zone radiation models using an Angstrom-Prescott equation and artificial neural networks. Energy 33, 1115–1127.

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. 3rd ed. p. 630.

World Meteorological Organization, 1981. Meteorological Aspects of the Utilization of Solar Radiation as an Energy Source. World Meteorological Organization Technical Note No. 172, WMO-No. 557, Geneva, p. 298.

Zhang, F., Zhou, X., Zhang, H., Peng, X., Wang, Z., 2014. On the relationship between direct and anisotropic diffuse radiation. Infrared Phys. Technol. 65, 5–8.

Zhao, N., Zeng, X., Han, S., 2013. Solar radiation estimation using sunshine hour and air pollution index in China. Energy Convers. Manag. 76, 846–851.