Energy Conversion and Management 126 (2016) 826–836

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Review

Assessment of ANN and SVM models for estimating normal direct irradiation (H_b)

Cícero Manoel dos Santos[a], João Francisco Escobedo[a,*], Érico Tadao Teramoto[b], Silvia Helena Modenese Gorla da Silva[b]

[a] *Rural Engineering Department, FCA/UNESP, Botucatu, São Paulo, Brazil*
[b] *Department of Fishing Engineering, São Paulo State University, Experimental Campus of Registro, Brazil*

ARTICLE INFO

Article history:
Received 7 June 2016
Received in revised form 12 August 2016
Accepted 13 August 2016
Available online 29 August 2016

Keywords:
Solar irradiation
Statistical models
Machine learning techniques
Meteorological variables

ABSTRACT

This study evaluates the estimation of hourly and daily normal direct irradiation (H_b) using machine learning techniques (ML): Artificial Neural Network (ANN) and Support Vector Machine (SVM). Time series of different meteorological variables measured over thirteen years in Botucatu were used for training and validating ANN and SVM. Seven different sets of input variables were tested and evaluated, which were chosen based on statistical models reported in the literature. Relative Mean Bias Error (rMBE), Relative Root Mean Square Error (rRMSE), determination coefficient (R^2) and "d" Willmott index were used to evaluate ANN and SVM models. When compared to statistical models which use the same set of input variables (R^2 between 0.22 and 0.78), ANN and SVM show higher values of R^2 (hourly models between 0.52 and 0.88; daily models between 0.42 and 0.91). Considering the input variables, atmospheric transmissivity of global radiation (kt), integrated solar constant (H_{sc}) and insolation ratio (n/N, n is sunshine duration and N is photoperiod) were the most relevant in ANN and SVM models. The rMBE and rRMSE values in the two time partitions of SVM models are lower than those obtained with ANN. Hourly ANN and SVM models have higher rRMSE values than daily models. Optimal performance with hourly models was obtained with $ANN4^h$ (rMBE = 12.24%, rRMSE = 23.99% and "d" = 0.96) and $SVM4^h$ (rMBE = 1.75%, rRMSE = 20.10% and "d" = 0.96). Optimal performance with daily models was obtained with $ANN2^d$ (rMBE = −3.09%, rRMSE = 18.95% and "d" = 0.97) and $SVM2^d$ (rMBE = 0.60%, rRMSE = 19.39% and "d" = 0.97). ANN and SVM models improved H_b estimations as compared with other results from the literature. SVM has better performance than ANN to estimate H_b, and it should be the first option of choice.

Contents

[*] Corresponding author at: Faculdade de Ciências Agronômicas (FCA/UNESP), Laboratório de Radiometria Solar, Departamento de Engenharia Rural, Fazenda Lageado, Rua José Barbosa de Barros, n° 1780, Botucatu, SP 18.610-307, Brazil.
E-mail address: escobedo@fca.unesp.br (J.F. Escobedo).

http://dx.doi.org/10.1016/j.enconman.2016.08.020
0196-8904/

Nomenclature

ANN	Artificial Neural Network
CosZ	cosine of the zenith angle
“d”	Willmott index
H_{sc}	integrated solar constant at the top of the atmosphere (MJ m^{-2})
H_0	solar irradiation at the top of the atmosphere (MJ m^{-2})
H_b	Normal direct irradiation (MJ m^{-2})
I_0	solar constant (W m^{-2})
kt_b	transmitted fraction
kt	atmospheric transmissivity of global irradiation
k_D	diffuse fraction
ML	machine learning
m_r	optical mass of the air
MLP	MultiLayer Perceptron
N	photoperiod
n	Sunshine duration
RBF	Radial Basis Function
rMBE	Relative Mean Bias Error (%)
rRMSE	Relative Root Mean Square Error (%)
R^2	Determination coefficient
r'	insolation ratio (n/N)
SACZ	South Atlantic Convergence Zone
SVM	Support Vector Machine
T_{max}	maximum air temperature (°C)
T_{min}	minimum air temperature (°C)
w_{ij}	hidden layer with linked weights
WEKA	Waikato Environment for Knowledge Analysis
x_{ij}	input layer
y_i	output layer
φ	latitude (in degrees)
δ	solar declination angle (in degrees)
ωs	half day length (in degrees)
C, γ and ε	RBF parameters
θ_i	bias of neuron *i*

Subscripts

h	hourly
d	daily

1. Introduction

Recent studies have shown the importance of reliable measurements of normal direct irradiation (H_b) for applications, such as calibration of satellites, study on thermal comfort and natural lighting of buildings and simulation of performance in concentrated solar technologies (CST) [1–3]. Because of competitive prices in the energy market and cost reduction in recent years, concentrated and photovoltaic solar energy has become the major alternative energy sources for the future. Therefore, long-term H_b measures are essential for financial planning, performance analysis and development of solar power system technologies.

Although there are sets of data on solar radiation and various solarimetric maps worldwide, they are generally not detailed enough to be used for determining solar energy available in small areas [4]. In addition, solar maps and H_b measurements are not readily available in most places of the world. H_b measurements are obtained through pyrheliometers. The major problem found for obtaining H_b measurements is the high cost for acquiring the sensor, tracking systems, adjacent devices and the need for regular maintenance to ensure measurement accuracy [5]. Without these sensors, studies have been conducted to recover historical series and provide estimates in places where measurements are not performed or not readily available. Estimates are typically obtained using models based on radiative transfer (sophisticated computer codes) or using decomposition models. Models based on radiative transfer are considered complex, require availability of a large number of input variables, and are useful only in clear-sky conditions [1]. Decomposition models are empirical in nature, locally adjusted and calculate H_b. They correlate the fractions of H_b components (transmitted fraction, kt_b), atmospheric transmissivity (kt), diffuse fraction (k_D) or insolation ratio (r' = n/N, n is duration of solar brightness and N is the photoperiod) [6].

When many input variables are used, decomposition models become complicated, time consuming and using multiple linear regression becomes inadequate [7]. However, these designs have inherent uncertainties that make their use limited. Therefore, estimating H_b is not so simple. Thus, new approaches are needed when the existing ones become limited or inefficient for some situations.

In the last 20 years, machine learning techniques (ML) have been tested and used to estimate solar radiation and have shown to be a good tool [8–11]. Using ANN and comparing it with empirical models, Soares et al. [8] estimated diffuse solar radiation in São Paulo city. The authors obtained RMSE from 0.193 to 0.121 MJ m^{-2}, with better performance for ANN than the empirical models. In Ghardaïa (Argelia), Belaid and Mellit [9] estimated daily and monthly global solar radiation using SVM. The authors combined different input variables amounting to 42 models. The results revealed good concordance among measures, and estimates with RMSE ranged from 2.727 to 2.807 MJ m^{-2}. There are several ML models and the Artificial Neural Network (ANN) and Support Vector Machine (SVM) are the most widely used [12–16]. Yadav and Chandel [13] reviewed the main studies in the literature using ANN to estimate solar radiation. The study reveals that ANNs estimate solar radiation more accurately than conventional methods. Raghavendra and Deka [14] reviewed SVM studies on hydrology and pointed out many examples of successful SVMs applications for modeling different hydrological processes. In Isfahan city, Mohammadi et al. [16] used support vector regression (SVR) to estimate global solar radiation using n and N as input variables. The authors compared the performance of SVR with that of empirical models and obtained RMSE = 2.004 MJ m^{-2} for daily estimates and RMSE = 0.450 MJ m^{-2} for monthly estimates with SVR, which shows better performance for ML. Considering both ML and SVM, the latter has better performance solving classification and regression problems due to its better generalization ability [9,17]. There are several studies assessing estimates of solar radiation data using ML and most of them analyzed estimates of daily global radiation [18,19]. Applying Support Vector Machine (SVM), Chen et al. [18] estimated global solar radiation for three places in Liaoning, China.

The authors used seven different combinations for input variables of SVM and compared the results with empirical models. They highlighted the supremacy of SVM (RMSE ranging from 1.789 and 2.380 MJ m^{-2}) over empirical models (RMSE ranging from 1.975 to 2.729 MJ m^{-2}).

Despite successful application in many areas, studies related to application of SVM to estimate H_b are rare, and those using ANN are also few [20–22]. In South America, including Brazil, measured data of H_b are scare or inexistent. Therefore, estimating H_b is essential for recovery and development of reliable historical series. Applications of ANN and especially SVM on renewable energy areas have been minimal [23]. Thus, the main aims of this study are: 1 – to analyze stability, accuracy and to exploit the potential of ANN and SVM to estimate H_b compared to classical statistical models; 2 – to investigate the key input variables in H_b modeling; 3 – to compare and indicate the best technique for H_b modeling. For a more extensive analysis, the models are evaluated in hourly and daily partitions.

The results serve as a case study because the Brazilian government has increased interest in using new renewable energy to meet current energy matrix (hydroelectric power plant). Brazil is the fifth largest country in the world; only a few research centers perform routine H_b measurements, however. Thus, there is a need for new methods to map H_b of a location and expand it to the whole territory. The study is divided into sections. Section 2 briefly shows the statistical models, ANN and SVM techniques used, the input variables and the validation indexes. The place and data set characteristic of modeling are given in Section 3. Section 4 discusses the results. The study is completed in Section 5.

2. Description of the methods used

The first part corresponds to generation of statistical models to estimate H_b, the second one, to estimate H_b using ANN, and the third one, to estimate H_b using SVM. Some concepts of each methodology are explained, and for more details, reading studies taken as reference is recommended. The location chosen for the case study is Botucatu, a city in the inner State of São Paulo/Brazil. Botucatu is the unique city in the Midwestern region of São Paulo State which has a 10 year-measurement H_b database. Because of technical problems in the pyrheliometer, monitoring of H_b ceased in 2009. In this case, the estimating of H_b is important to the completion of the temporal series in Botucatu. To estimate H_b, organizing the data set and choosing the appropriate algorithm are highly important. The description of the site and data used is presented in Section 3.

2.1. Statistical models

The statistical models were developed according to the same input variables of ANN and SVM models (Table 1), and were separated into hourly and daily partitions. The first model, in hourly partition, uses the relationships between H_b and H_G, the second model correlates H_b with kt and the third model correlates H_b with kt and H_{sc}. The statistical models in the daily partition follow the same structure of the hourly models, altering the input variables to find the best correlations to estimate H_b. The statistical models were adjusted by polynomial regression. Following that sequence of input variables, if the statistical models show determination coefficient (R^2) lower than that of ANN and SVM models of the same structure, they will be rejected and the results will be developed as a function of estimates of H_b using ANN and SVM.

2.2. Artificial neural network

ANN is a computer system for processing information that consists of an interconnected group of artificial neurons based on the structure, processing method and the brain learning ability [24]. ANN is able of storing knowledge and understanding the complex non-linear relationship between output and input data, covering regression problems, forecasting models and other applications in different fields [25–27].

Table 1
Set of input variables that define ANN and SVM models.

Partition		ANN	SVM	Set of input variables
Hourly	H_b^h	$ANN1^h$	$SVM1^h$	H_G
		$ANN2^h$	$SVM2^h$	kt
		$ANN3^h$	$SVM3^h$	kt and H_{sc}
		$ANN4^h$	$SVM4^h$	kt and m_r
		$ANN5^h$	$SVM5^h$	Kt and CosZ
		$ANN6^h$	$SVM6^h$	kt, H_{sc}, m_r and CosZ
Daily	H_b^d	$ANN1^d$	$SVM1^d$	H_G
		$ANN2^d$	$SVM2^d$	kt
		$ANN3^d$	$SVM3^d$	kt and H_{sc}^d
		$ANN4^d$	$SVM4^d$	kt and r′
		$ANN5^d$	$SVM5^d$	kt, H_{sc}^d and r′
		$ANN6^d$	$SVM6^d$	T_{max} and T_{min}

H_b is the normal direct irradiation (MJ m^{-2}); kt is atmospheric transmissivity; H_{sc}: integrated solar constant at the top of the atmosphere (4.921 MJ m^{-2}); m_r is the optical mass of the air; CosZ is the cosine of the zenith angle; H_{SC}^d is the daily direct irradiation at the top of the atmosphere (MJ m^{-2}), r′ is the insolation ratio (n/N), T_{max} is the maximum air temperature, T_{min} is the minimum air temperature.

The ANN adopted in this paper is the MultiLayer Perceptron (MLP). MLP is an information processing system massively parallel and distributed, and applied successfully to a model for many non-linear problems [28,29]. The basic structure of the MLP is an input layer (x_{ij}), hidden layer with linked weights (w_{ij}), and an output layer (y_i) [30]:

$$y_i = \sum_{j=1}^{\eta} w_{i,j} x_{i,j} + \theta_i \tag{1}$$

where $x_{i,j}$ is the input signal from the *j-th* neuron (for input layer), $w_{i,j}$ is the weight of the direct connection of neuron j to neuron i (in the hidden layer), and θ_i is the bias of neuron i. The output of neurons is calculated by applying an activation function. The activation function used is typically standard sigmoid (Eq. (2)).

$$f(x) = \frac{1}{[1 + \exp(-x)]} \tag{2}$$

In this study, MLP was trained using the Backpropagation training algorithm and the term momentum [24,31]. The weight adjustment at the iteration depends on the learning rate and momentum. The learning rate during each iteration controls the size of weight and bias changes, while the momentum helps the search for the global minimum on the error surface, preventing the system from converging to a local minimum or saddle point [32].

2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning method derived from the statistical learning theory introduced by [17]. SVM has gained prominence in many fields of knowledge to solve complex problems of pattern recognition, classification, regression analysis and forecasting [33–35].

The performance and learning capacity of SVM in regression are attributable to the use of the Kernel function set, which diagrams the information to a higher dimensional space [36], which makes the SVM a feasible choice to address several solar radiation studies on non-linear nature [37]. In this study, the Kernel Radial Basis Function (RBF) is considered for SVM regression. The best estimate function $f(\chi)$ may be expressed as an expansion of support vectors [38]:

$$f(\boldsymbol{\chi}) = \sum_{i=1}^{T} \boldsymbol{\beta}_i k(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) + b \tag{3}$$

where $k(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) = \exp(-0.5 \times \|\boldsymbol{\chi} - \boldsymbol{\chi}_i\|^2/\sigma^2)$ is RBF, the value of the kernel function $k(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)$ equals to the intern product of two vectors $\boldsymbol{\chi}_i$ and $\boldsymbol{\chi}_j$ in the feature space $\boldsymbol{\varphi}(\boldsymbol{\chi}_i)$ and $\varphi(\boldsymbol{\chi}_j)$ respectively; the multiplicators $\boldsymbol{\beta}_i \in [-C, C]$, for $i = 1, \ldots, T$ are the solutions for the problem of double optimization in SVM regression. The points $\boldsymbol{\chi}_i$ with multiples different from zero $\boldsymbol{\beta}_i$ are called Support Vectors (SVs). The scalar b is estimated by minimizing the sum of the empirical risk following by introduction of positive slack variables [33].

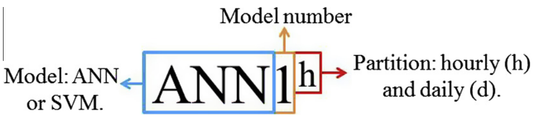
To allow greater flexibility in the application of RBF, it is necessary to properly adjust the parameters *C*, γ and ε. Parameters *C* and γ are mutually dependent on each other, low *C* values produce machine learning with poor approach, and high *C* values generate more complex learning machine [14]. In this study, γ and C values are tested by trial method and those with the best accuracy for cross-validation are chosen. Parameter ε is used to adjust the training data.

The SVM used is the integrated LibSVM compilation for classification of support vectors, regression and distribution estimate [39,40]. A brief introduction to the theory of SVM is presented in [18].

2.4. Software and models

The ANN models were trained and validated using Waikato Environment for Knowledge Analysis (WEKA). WEKA is a set of ML algorithms for data mining tasks [41]. Providing the user with a Java programming, WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. WEKA has the option of choosing the percentage of data for training and validation, defined as Percentage Split. The choice of data from this tool is at random, with no tendency in modeling. Seventy percent of the series were used for training and 30% of them for validation and testing. Algorithm J48 is used to classify the most relevant input variables to estimate H_b. The algorithm J48 (implemented in WEKA) is widely used to construct a decision tree [42,43]. It is used for rules of classification and represents the knowledge based on the tree. Moreover, it consists of a great number of branches, a root, some knots and leaves. A branch is a chain of ganglions from the root to the leaves, and each knot involves one variable. The occurrence of one variable in one tree supplies information on the importance of the associated variable. Therefore, selection of the most relevant variables to estimate H_b is performed using the decision tree method.

The correlations between H_b and other input variables are shown in Table 1. The input data for ANN and SVM models consist of independent and dependent variables. Different combinations with input variables were formed in hourly and daily partitions. The input variables were selected because of their correlation with H_b and they are more easily monitored and available in stations than other atmospheric parameters (aerosols and water vapor, for example). In hourly and daily partitions: H_G choice is justified because H_b at the horizontal is one of their components. Selection of kt and the insolation rate (r′ = n/N) is explained by the similarity to Liu & Jordan and Ångström methodologies, respectively. H_{sc} is justified due to its relationship with H_b, and m_r because it is a variable which reduces the incidence of H_b. The cosZ is justified by the increase in H_b with elevation of the zenithal angle and vice versa. T_{max} and T_{min} input variables are evaluated to test them regarding the estimate of H_b, as well because of simplicity of measurement and low cost of the devices used. The models trained with Multi-Layer Perceptron and with Support Vector Machine are represented as ANN and SVM, respectively. Abbreviations of models:



Solar irradiation at the top of the atmosphere (H_0) was calculated according to the equations described in [44,45], which is a function of solar constant (I_0), Julian day, latitude (φ, in degrees), solar declination angle (δ, in degrees) and half day length (ω_s, in degrees). The photoperiod (N) was calculated according to Eq. (4):

$$N = \frac{2}{15} \times \omega_s \tag{4}$$

Daily direct irradiation at the top of atmosphere (H^d_{sc}, MJ m^{-2}) is obtained by multiplying the integrated solar constant (H_{sc} = 4.921 - MJ m^{-2}) by the photoperiod (Eq. (5)):

$$H^d_{sc} = 4.921 \times N \tag{5}$$

Optical relative mass (m_r) was calculated according to [46].

2.5. Validation and evaluation of models

Various statistical indexes can be used to evaluate the performance of models to estimate H_b. Some indexes are used here to evaluate ANN and SVM models: Relative Mean Bias Error (rMBE), Relative Root Mean Square Error (rRMSE), determination coefficient (R^2) and "d" Willmott index [47]. The definitions of those indexes are given by the following equations:

$$\mathrm{rMBE}(\%) = 100 \times \frac{\frac{\sum_{i=1}^{X}(H_E - H_M)}{X}}{\overline{X}} \tag{6}$$

$$\mathrm{rRMSE}(\%) = 100 \times \frac{\left[\frac{\sum_{i=1}^{X}(H_E - H_M)^2}{X}\right]^{\frac{1}{2}}}{\overline{X}} \tag{7}$$

$$R^2 = \left(\frac{\sum_{i=1}^{X}(H_M - \overline{H}_M)^2 \times \sum_{i=1}^{X}(H_E - \overline{H}_E)^2}{\sqrt{\sum_{i=1}^{X}(H_M - \overline{H}_M)^2} \times \sqrt{\sum_{i=1}^{X}(H_E - \overline{H}_E)^2}}\right)^2 \tag{8}$$

$$d = 1 - \frac{\sum_{i=1}^{X}(H_E - H_M)^2}{\sum_{i=1}^{X}(|H'_E| + |H'_M|)^2} \tag{9}$$

where H_E represents the estimated values, H_M the measured values, $|H'_E|$ the absolute value of the $H_M - \overline{H}_M$ and $\overline{H}_E$ difference, in which $\overline{H}_M$ represents the average of H_M, $|H'_M|$ represents the absolute value of the $H_M - \overline{H}_M$ difference and $\overline{H}_E$ the average of the estimated values. $\overline{X}(= \frac{1}{x}\sum_{i=1}^{x} H_M)$ is the average value of the measurement and *x* is the number of observations. The rMBE index describes the average trend of estimated values to overestimate (positive values) or underestimate (negative values) the measures. The optimal rMBE value is 0. rRMSE and R^2 indexes are often used. The optimal rRMSE value is 0. Good models should have low rRMSE and rMBE values. The R^2 value ranges from 0 to 1 and the higher its value, the better the model fits. The Willmott "d" concordance index indicates how close the estimates are from the measures of the comparison line 1:1. The "d" value equal to 1 corresponds to a perfect match. To find out which model can have a consistently high performance, Gueymard and Ruiz-Arias [48] elaborated a set of criteria for evaluation of H_b models. Different rRMSE intervals are defined to evaluate the accuracy of the models [49,50]:

Excellent if rRMSE < 10%;
Good if 10% ⩽ rRMSE < 20%;
Fair if 20% ⩽ rRMSE < 30%;
Poor if rRMSE ⩾ 30%.

In this work, these intervals serve as parameters to assess the accuracy of models to estimate H_b and define the most recommended one. This work considered low errors for |rMBE| < 10%.

3. Study site and database

Data used in this study were measured in the Radiometric Station located at the College of Agricultural Sciences in Botucatu - FCA/UNESP (22.85°S; 48.45°W and 786 m altitude) (Fig. 1). The region has high altitude gradient between 400 and 500 m in the lower region (peripheral depression) and between 700 and 900 in the mountain region (Western Highlands). With Savanna and Atlantic Forest biome, it has warm temperate (mesothermal), hot and wet summer with high precipitation and dry winter [47]. With annual average air temperature of 20.46 ± 2.21 °C, the hottest month is February (23.216 ± 1.20 °C) and the coldest month is July (17.16 ± 1.33 °C). Relative humidity ranges from 62.61 ± 8.88% (August) to 76.26 ± 8.24% (February). Accumulated annual average precipitation is 1,494.10 mm. The rainiest season is from October to March and the driest season is from April to September. During the rainy season, precipitation is mainly caused by the South Atlantic Convergence Zone (SACZ) and frontal systems. In the dry season, rainfall originates from the meeting of cold and dry air masses coming from the southern region with warm and moist masses from the southeastern region [51].

H_b was measured by an Eppley NIP pyrheliometer coupled to an Eppley ST3 solar tracker, H_G by Eppley PSP pyranometer, and sunshine duration by a conventional Campbell-Stokes sunshine recorder. Measurements of maximum and minimum air temperatures were performed using a mercury thermometer and alcohol thermometer, respectively. All meteorological variables were measured during the period from February 1996 to December 2008. The sensors are daily checked for replacement whenever necessary. Irradiance measurements were taken every minute, storing the average every 1 min. Irradiance ($W\,m^{-2}$) is integrated at specific time intervals to obtain radiation ($MJ\,m^{-2}$) from programs for radiation analysis [52]. Radiometers are annually assessed for different sky covers by the comparative method suggested by the World Meteorological Organization [53]. As there is no definitive, ideal or widely accepted procedure for better quality control of irradiation data, each research center typically adopts its own method, which implies that some may be more accurate than others [48].

Table 2
Determination coefficient (R^2) obtained using ANN and SVM models.

Hourly				Daily			
ANN	R^2	SVM	R^2	ANN	R^2	SVM	R^2
ANN1^h	**0.53**	**SVM1^h**	**0.52**	**ANN1^d**	**0.49**	**SVM1^d**	**0.55**
ANN2^h	0.87	SVM2^h	0.87	ANN2^d	0.87	SVM2^d	0.87
ANN3^h	0.87	SVM3^h	0.87	ANN3^d	0.88	SVM3^d	0.86
ANN4^h	0.88	SVM4^h	0.87	ANN4^d	0.88	SVM4^d	0.89
ANN5^h	0.88	SVM5^h	0.87	ANN5^d	0.91	SVM5^d	0.91
ANN6^h	0.88	SVM6^h	0.87	**ANN6^d**	**0.42**	**SVM6^d**	**0.43**

4. Results and discussion

In ANN models, the following values were considered: learning rate = 0.3; momentum = 0.2 and number of iterations = 500. Hidden layers were tested and ranged from 1 to 10, but the standard value of WEKA was adopted by the best fit found. In WEKA, the standard of hidden layers is defined as "α" = [(Input variables + classes)/2]. In the SVM training, selection and proper use of the Kernel function have great accuracy on the modeled data and on SVM models [54]. The three RBF parameter set were: C, γ and ε. In their selection, the ε value was set at 0.005 and many assays were performed with different C and γ combinations. So, the best values were: $C = 50$, $\gamma = 0.2$ and $\varepsilon = 0.005$. Two combinations of RBF parameters suggested for H_G were tested: $C = 100$; $\gamma = 0.3$; $\varepsilon = 0.001$ [37] and $C = 400$; $\gamma = 0.01$; $\varepsilon = 0.4$ [16], but the R^2 values generated for models with these parameters were lower than those obtained in this study for H_b (Table 2).

Models with the lowest R^2 Willmott values are highlighted in bold. The smallest R^2 value (0.42) was obtained using the ANN6^d and the highest R^2 values (0.91) were obtained using two models (ANN5^h and SVM5^d). In general, the similarity in the R^2 values between ANN and SVM and the efficiency in modeling gain prominence, especially when parameters are properly adjusted and good measures are used [27].

Fig. 2(a, b, c and d) shows the correlations $H_b \times H_G$ and $H_b \times k_t$ in hourly and daily time partition with adjusted equations of their correlations. Data were analyzed through a dispersion curve and after exclusion of inconsistent values. In the adjusted polynomial equation, values of R^2 were low when compared to those from the ML models (Table 3). Only ANN and SVM models will be compared with each other due to low R^2 values shown by the statistical models.

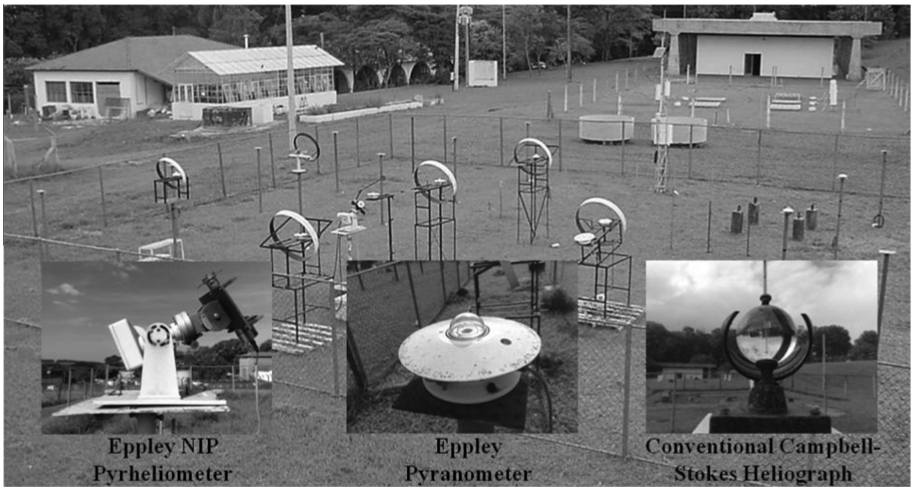


Fig. 1. Radiometric Station located at the College of Agricultural Sciences from Botucatu - FCA/UNESP.

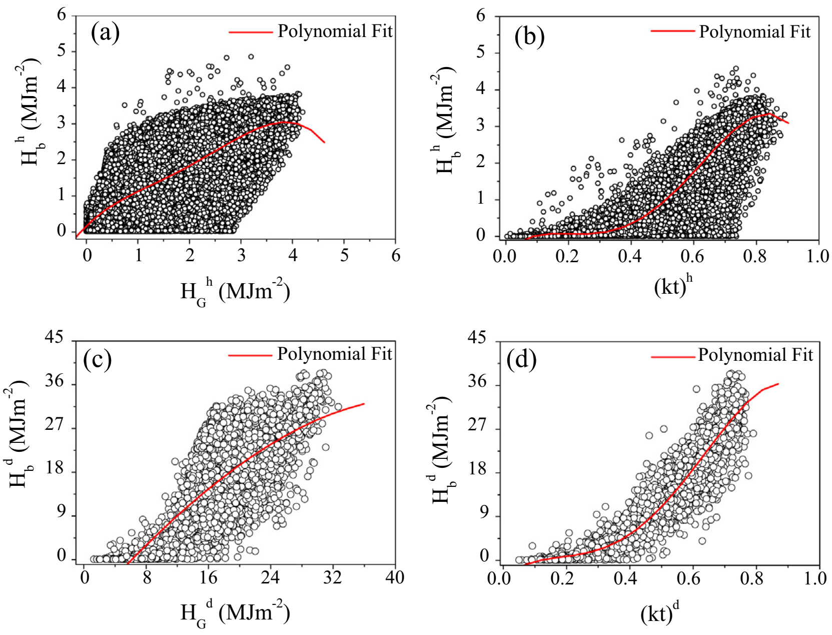


Fig. 2. Hourly and daily adjusted statistical models: (a) $H_b^h \times H_G^h$ correlation, (b) $H_b^h \times k_t^h$ correlation, (c) $H_b^d \times H_G^d$ correlation and (d) $H_b^d \times k_t^d$ correlation.

Table 3
Adjusted statistical models.

Figs.	Statistical models	R^2
Fig. 2a	$H_b^h = 0.17 + 1.40 \times H_G^h - 0.69 \times \left(H_G^h\right)^2 + 0.27 \times \left(H_G^h\right)^3 - 0.04 \times \left(H_G^h\right)^4$	0.22
Fig. 2b	$H_b^h = -0.44 + 8.60 \times kt^h - 48.60 \times (kt^h)^2 + 105.55 \times (kt^h)^3 - 63.66 \times (kt^h)^4$	0.77
Fig. 2c	$H_b^d = -10.60 + 1.80 \times H_G^d - 0.01 \times (H_G^d)^2 - 1.45E - 4 \times \left(H_G^d\right)^3$	0.29
Fig. 2d	$H_b^d = -3.64 + 56.65 \times kt^d - 290.77 \times (kt^d)^2 + 677.60 \times (kt^d)^3 - 410.93 \times (kt^d)^4$	0.78

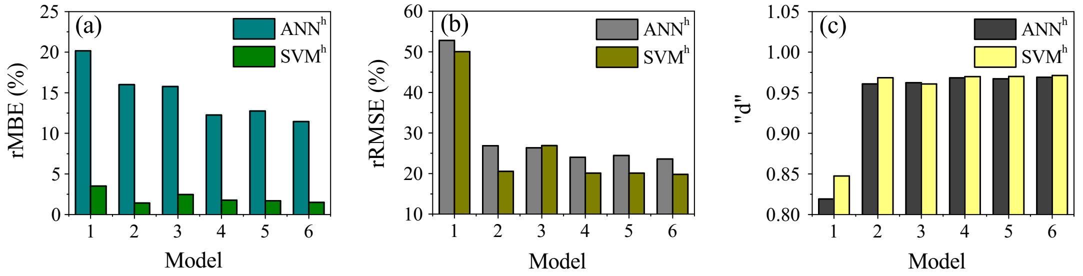


Fig. 3. Validation indexes of ANN^h and SVM^h models to estimate H_b^h: (a) rMBE, (b) rRMSE and (c) “d” Willmott index.

In addition to checking the best machine learning technique, this study also identified the input climate variables that are important for determination of H_b. For that purpose, the WEKA J48 algorithm was used as classifier, as the most relevant input variables kt and H_{sc} were obtained in hourly models; kt and insolation ratio (r′ = n/N) were obtained in daily models. The best R^2 values are found for models with the addition of these variables. kt and m_r as the most relevant input variables in the H_b modeling are highlighted in the study by [7].

4.1. Performance analysis of hourly models

Because of lack of studies using ML to estimate H_b [13], the results in this study are limited to a few comparisons with those from the literature, but there is a need for exploring and validating new methodologies to estimate H_b to meet the global demand for solarimetric information for different applications, including calibration of satellites, solarimetric mapping and retrieving historical series.

In assessing the results with models (ANN^h and SVM^h), validation indexes (rMBE, rRMSE and d) were used (Fig. 3a, b, c). All models overestimated the measures (rMBE > 0). The ANN^h models estimated the measures (rMBE > 10.0%). The rMBE values obtained with ANN^h are on average 12.68% higher than those obtained with SVM. The rRMSE and “d” results obtained with SVM models are lower and higher, respectively than those generated using ANN models. SVM1^h (rRMSE = 50.05%) and ANN1^h (rRMSE = 52.79%) showed poor accuracy in estimating H_b^h. The rMBE values (20.16% with ANN1^h and 3.51% with SVM1^h) and “d” (0.82 with ANN1^h and 0.85 with SVM1^h) were the highest and the lowest in the hourly models, respectively. The result proves that only H_G^h as input variable is not recommended for estimating H_b^h. Temporal variation of H_b as compared to H_G is higher [55], therefore, the more difficult to model, the lower the correlation between the two components. So, new input variables are presumably necessary to generate better models and estimates for H_b [56]. Therefore, the following models of new variables were added. With the addition of kt^h in $ANN2^h$ and $SVM2^h$ models, there was an improvement in rMBE, rRMSE and "d" values. Thus, $ANN2^h$ and $SVM2^h$ models estimated H_b^h with rRMSE = 26.85% and rRMSE = 20.55%, respectively, and overestimated measures with rMBE = 16.00% and rMBE = 1.40% for $ANN2^h$ and $SVM2^h$, respectively. $SVM2^h$ has better "d" (0.97) than $ANN2^h$ (d = 0.96). The use of kt^h as input variable is interesting because it is indicative of cloudy sky conditions, eliminates astronomical effects and points out climate effects.

In $ANN3^h$ and $SVM3^h$ models, the addition of H_{sc} in the set of input variables did not improve the performance of the models. The insertion of m_r ($ANN4^h$) improved rRMSE by ≈2.86% when compared to $ANN2^h$ and by ≈2.35% with respect to $ANN3^h$. $ANN5^h$ and $ANN6^h$ models have results similar to $ANN4^h$. The best performance among network models is for $ANN6^h$: rMBE = 11.45%, rRMSE = 23.58% and d = 0.97. With the inclusion of new variables, there was a gradual improvement in neural network models. However, no significant difference was found in rRMSE and "d" values for $ANN4^h$–$ANN6^h$ models. Except for $ANN1^h$, combinations of input variables of neural network models may be used to estimate H_b^h.

In SVM models ($SVM2^h$ – $SVM6^h$), the addition of new input variables did not significantly improve the estimate of H_b^h. In these models, rMBE values were between 1.40 and 3.51%; rRMSE between 19.78 and 20.55%; d > 0.96. Therefore, SVM shows better stability than ANN in modeling H_b^h. Some modifications in the parameters of the RBF function (C and γ) were performed to determine whether estimation for models improved when performed separately. Since there was no gain in modeling, the values already highlighted remain fixed. Thus, stability occurs when the value of parameter C is properly defined and then the discrepant values of parameters are easily supplied. Therefore, the results confirm that SVM is able to produce accurate results [57], has higher generalization capacity and potential to track historical data to improve future forecasting series [58].

The results of this study are similar to those found by [10], who used ANN to estimate H_b^h by satellite image and obtained rRMSE = 26.10% and rMBE = −6.01%. Linarez-Rodriguez et al. [11] showed that in studies of the literature, H_b^h estimates from satellites have average rRMSE value ~35.00%. From a proposed method, Fernández-Peruchena et al. [59] generated a climate series of H_b with values every 1 min and found average error >30.0%. Generating synthetic H_b series every 5 min, Grantham et al. [60], obtained rMBE = −0.40% and rRMSE = 16.30%. The difference between local results and those above is related to the methodology adopted. The method used by the authors shows sky conditions at intervals with increment of 0.01, which is different from the local methodology. The local results are better than those obtained by Polo et al. [61], who reported rRMSE = 31.0% in the generation of H_b series every 10 min from a given time series. Linarez-Rodriguez et al. [11] estimated H_b^h by satellite using five ANN models and obtained rRMSE between 24.23 and 37.60%. The finding shows that local models and the used variables had better performance in estimating H_b^h. That difference is mainly due to larger errors expected when satellite products are used. The rMBE values found by [11] resemble those locally obtained.

Box graphs show the errors among measurements and H_b^h estimates using ANN^h (Fig. 4a) and SVM^h models (Fig. 4b). Inside each box, the center mark shows the average of all error values. By default, the box is determined by 25th and 75th percentiles. Whiskers are determined by 5th and 95th percentiles. Except for $ANN1^h$ and $SVM1^h$ models, the others have narrow box, i.e., the estimates are more reliable. These results agree with the performance of models analyzed using validation indexes.

Dispersion between measured and estimated H_b^h using models that have lower and higher accuracy is shown in Fig. 5a–d. $ANN1^h$ (Fig. 5a) and $SVM1^h$ models (Fig. 5b) had the highest dispersion. The high dispersion found for these models is because climatic effects of H_b^h are greater than astronomical effects. That is, temporal variability of H_b^h in a cloudy or turbid atmosphere is different from temporal variation of H_b^h under the same atmosphere. The increase in cloudiness generates a reduction in atmospheric transmission of H_b^h, but attenuation in H_G^h is not proportional to the attenuation by H_b^h transmission because of the conversion of direct irradiation into diffuse irradiation. Transition of atmosphere from partially cloudy to cloudy generates 100% reduction in H_b, while only 26% reduction in H_G [62]. In the absence of clouds, aerosols are the main responsible for reducing H_b [63]. Depending on the atmospheric turbidity, reduction in H_b by aerosols can vary from 30% to 100%, while in H_G is less than 10%. Due to low correlation between increase and reduction of radiation, H_G alone should not be directly applied to estimate H_b [2]. $ANN6^h$ (Fig. 5c) and $SVM6^h$ models (Fig. 5d) had the lowest dispersion and show good agreement. The curves of the other models are not shown, but they have behavior similar to that of the above models.

4.2. Performance analysis of daily models

The validation indexes (rMBE, rRMSE and d) of daily models show that estimation with SVM^d is better than estimation with ANN^d (Fig. 6a, b, c). Models, whose input variables are only H_G^d, $ANN1^d$ and $SVM1^d$ estimated the measures with rRMSE = 26.91% and rRMSE = 22.08%, respectively. With the inclusion of kt, $ANN2^d$ decreased by ≈7.96% the rRMSE value compared with $ANN1^d$. The

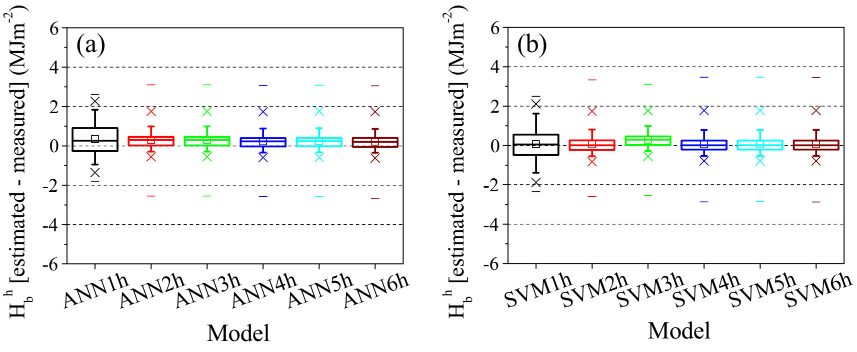


Fig. 4. Box plots of errors of fourteen hourly models: (a) ANN^h models and (b) SVM^h models.

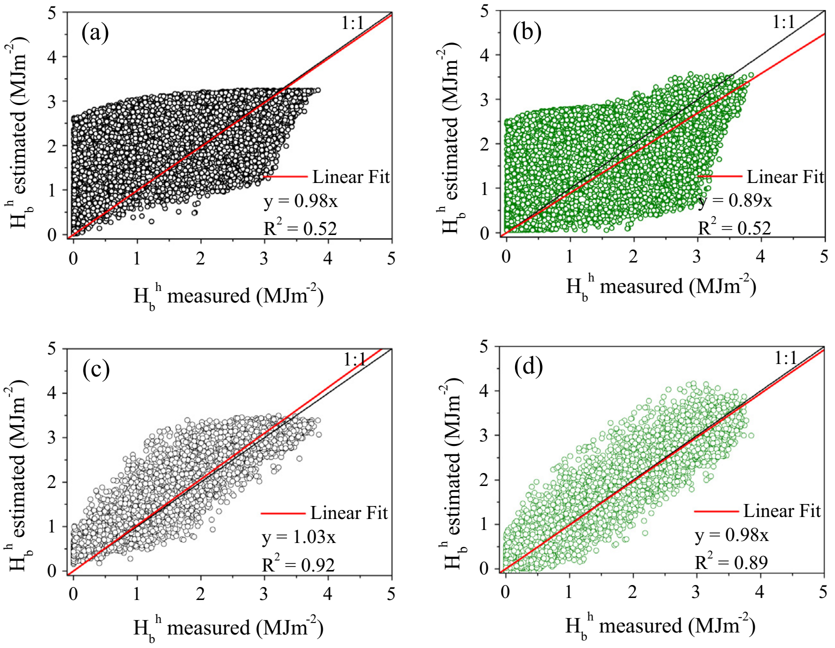


Fig. 5. Dispersion of data measured and estimated by ANN and SVM for H_b^h: (a) ANN1^h model, (b) SVM1^h model, (c) ANN6^h model and (d) SVM6^h model.

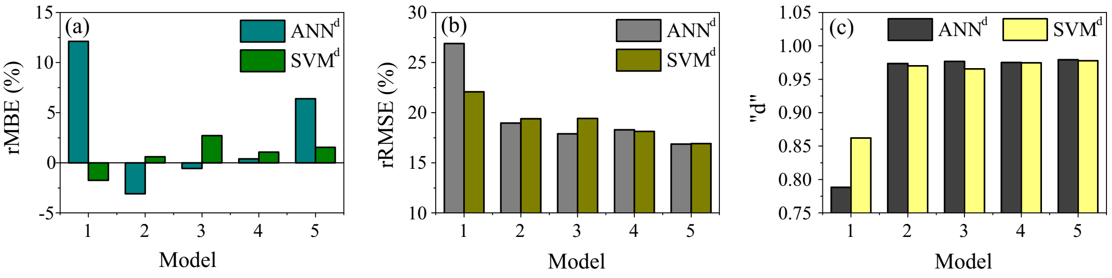


Fig. 6. Validation indexes of ANN and SVM models in estimating H_b^d: (a) rMBE, (b) rRMSE and (c) “d” Willmott index.

ANN3^d model did not improve estimation with addition of H_{sc}^d compared with ANN2^d. ANN4^d improved the estimate (rMBE = 0.38%, rRMSE = 18.30% and d = 0.98) with the addition of the insolation ratio variable. When ANN4^d is compared with ANN1^d, the rRMSE reduction was ≈8.30%, with ANN2^d, the reduction was ≈8.61%. The ANN5^d model shows rMBE = 6.39%, rRMSE = 16.85% and d = 0.98. The results obtained in the validation of ANN6^d and SVM6^d models were not shown here due to their high inaccuracy, which show the infeasibility of using these set of input variables. Therefore, the combination of ANN5^d variables is the most appropriate.

SVM models have smaller rMBE than ANN. SVM1^d is the only SVM model to underestimate (rMBE = −1.74%). The addition of new input variables improved accuracy of SVM models. SVM2^d–SVM5^d models have rMBE values ranging from 0.60 to 2.70% (mean = 1.48 ± 0.90%), rRMSE from 16.92% to 19.43% (mean 18.46 ± 1.19%) and d = 0.98. Although SVM input vectors are quite flexible, the replacement of these input variables with information from aerosols and water vapor would result in better estimations. This work did not consider aerosols and water vapor because these variables are not readily available.

ANN^d and SVM^d had better accuracy with insertion of other variables. In the case of ANN1^d and SVM1^d models, the explanation for the poor accuracy can be summarized as follows: clouds in the atmosphere result in reduced H_b^d values, which cause increase in daily diffuse irradiation (H_D^d). This dynamics in the atmosphere makes clear-sky days present high H_G^d and H_b^d values, while, in cloudy-sky days, H_b^d tends to zero and H_G^d is significantly reduced to lower values, but higher than H_b^d values due to the H_D^d component that raises H_G^d values. This dynamics makes the relationship between H_b and H_G be low, which affects modeling. Therefore, the results show that inclusion of more variables improves estimation of H_b^d with ANN and SVM.

The Box graphs show errors of measurements and H_b^h estimates with ANN^d (Fig. 7a) and SVM^d models (Fig. 7b). Inside each box, the center mark shows the average of all error values. By default, the box is determined by the 25th and 75th percentiles. Whiskers are determined by 5th and 95th percentiles. Except for ANN1^d and SVM1^d, the others have narrow box, i.e. estimates are more reliable. These results agree with the performance of models analyzed with validation indexes.

Some possible combinations have been considered to find a set of input variables and the most favorable learning technique in estimating H_b^d. Six models with different combinations of input variables have been established. The H_b^d values estimated using ANN and SVM, according to measures for the best and worst models, are presented in Fig. 8a-d. The dispersion with ANN1^d (Fig. 8a) and SVM1^d models (Fig. 8b) show low correlation between H_b^d and H_G^d. ANN4^d (Fig. 8c) and SVM5^d models (Fig. 8d) show that many points are along the ideal comparison line (1:1) as well as good

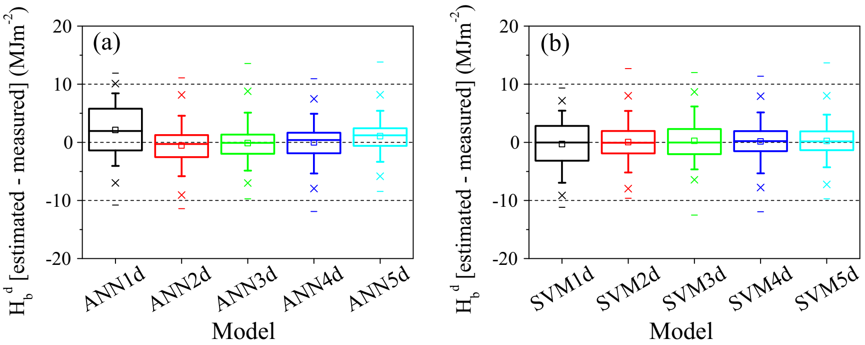


Fig. 7. Box plots of errors of sixteen daily models: (a) ANN^d models and (b) SVM^d models.

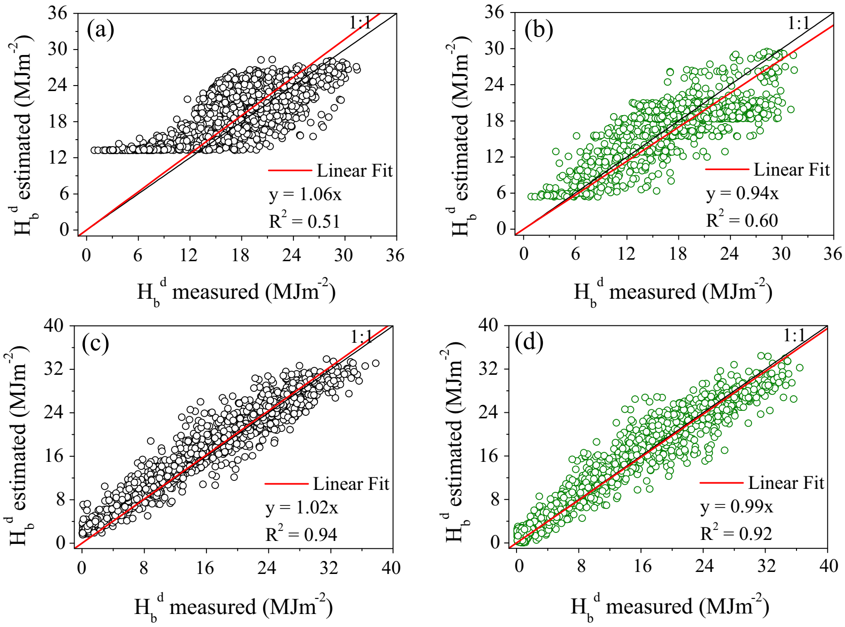


Fig. 8. Dispersion of measured data and those estimated by ANN and SVM for H_b^d: (a) ANN1^d model, (b) SVM1^d model, (c) ANN5^d model and (d) SVM5^d model.

agreement between measured and estimated values. The capacity of each model and technique selected to provide accurate estimates are related to the proper selection of input variables.

5. Conclusions

This study presents the use of ANN and SVM to estimate hourly and daily normal direct irradiation in Botucatu/SP. The following input variables were analyzed: H_G, kt, H_{sc}, cosZ, m_r, r′, T_{max} and T_{min}. In ANN and SVM, for training and validation of ANNs and SVM, different input variables were used to identify those presenting the best correlation with H_b. The performance of ANN and SVM is compared with that of statistical models using the same input variable. A total of 13-year data of high quality were used for training and validation of the ANN and SVM models, generation and validation of the statistical models. Estimate of H_b and assessment of methods are of great importance for application in the area of conversion of solar energy and in locating areas with potential for installation of solar power plants.

The statistical models presented determination coefficients (R^2) lower than those of ANN and SVM models. Therefore, the machine learning models have more generalization capability and ability to adapt different boundary conditions than the statistical models. Therefore, the statistical models in this study were considered as the third choice of option to estimate H_b.

In the estimation of H_b with SVM, the use of the following parameters is recommended: $C = 50$, $\gamma = 0.2$ and $\varepsilon = 0.005$ for RBF. Other kernel functions were of no interest in being evaluated. However, other studies may be considered in assessing the flexibility of other functions in the estimation of H_b. The most relevant input variables in the models were H_{sc}, kt and the insolation ratio.

In hourly estimates, the greatest errors are observed for ANN1^h: rMBE = 20.15%, rRMSE = 52.79% and d = 0.82; SVM1^h: rMBE = 3.51%, rRMSE = 26.84% and d = 0.85. The models with kt as input variables had: ANN2^h (rMBE = 16.00%, rRMSE = 26.84% and d = 0.96) and SVM2^h (rMBE = 1.40%, rRMSE = 20.54% and d = 0.97). In average, the models of ANN^h presented rMBE = 14.72 ± 3.25%, rRMSE = 29.66 ± 11.40 and d = 0.94 ± 0.06; SVM^h: rMBE = 2.05 ± 0.80%, rRMSE = 26.24 ± 11.97% and d = 0.95 ± 0.05. Except for ANN1^h and SVM1^h models, which were the models with the greatest errors, the models had: ANN^h (rMBE = 13.64 ± 2.09%, rRMSE = 25.04 ± 1.46% and d = 0.96 ± 0.003) and SVM^h (rMBE = 1.76 ± 0.42%, rRMSE = 21.45 ± 3.03% and d = 0.97 ± 0.004).

For the daily estimate, the results showed the worst estimate for the ANN1^d models: rMBE = 12.11%, rRMSE = 26.91% and d = 0.79; SVM1^h: rMBE = −1.74%, rRMSE = 22.08% and d = 0.86. The models ANN5^d and SVM5^d resulted in (rMBE = 6.39%, rRMSE = 16.85% and d = 0.98) and (rMBE = 1.55%, rRMSE = 16.92% and d = 0.98), respectively. In general, the ANN^d models presented average values of rMBE = 3.04 ± 6.15%, rRMSE = 19.78 ± 4.06 and d = 0.94 ± 0.08; SVM^d: rMBE = 0.83 ± 1.63%, rRMSE = 19.19 ± 1.91% and d = 0.95 ± 0.05.

Based on the results found in Botucatu, ML techniques were able to model different combinations of input data for all weather conditions. At first, the statistical models were discarded due to their high inaccuracy. As the comparison is performed between MLP and SVM, it was concluded that SVM has better performance than MLP because there are many other types of ANN, and the MLP evaluated in this study is only a kind of ANN. Therefore, SVM has better accuracy in estimating H_b than ANN and it is the first option of choice. The exploration and validation of new estimation methodologies to meet the demand of solarimetric information for different applications are of great importance in current studies.

Information and estimates of H_b (hourly and daily) are relevant to be used for comparison and calibration of satellite data, improving temporal and spatial resolution. The models can be applied to other locations where there are other measured variables recovered from historical series or dataset. Further studies should analyze other meteorological variables in H_b modeling. Analysis of techniques for shorter time intervals is interesting for solar projects. In order to determine accurate measurements, validating SVM in other locations is of great interest. The assessment of ANN and SVM in estimating other irradiations are proposed for future studies, as the demand for solarimetric information increases routinely.

Acknowledgments

The authors acknowledge the financial support provided by CAPES, CNPq and FAPESP.

References

[1] Gueymard CA. Direct solar transmittance and irradiance predictions with broadband models. Part I: detailed theoretical performance assessment. Sol Energy 2003;74:355–79.
[2] Law EW, Prasad AA, Kay M, Taylor RA. Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting – a review. Sol Energy 2014;108:287–307.
[3] AL-Rasheedi M, Gueymard CA, Ismail A, AL-Hajraf S. Solar resource assessment over Kuwait: validation of satellite-derived data and reanalysis modeling. In: EuroSun 2014/ISES conference proceedings.
[4] Badescu V, Gueymard CA, Cheval S, Oprea C, Baciu M, Dumitrscu A, et al. Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. Renew Energy 2013;55:85–103.
[5] Bertrand C, Vanderveken G, Journee M. Evaluation of decomposition models of various complexity to estimate the direct solar irradiance over Belgium. Renew Energy 2015;74:618–26.
[6] Gueymard CA, Ruiz-Arias JA. Performance of separation models to predict direct irradiance at high frequency: validation over arid areas. In: EuroSun 2014/ISES conference proceedings.
[7] Lopez G, Batlles F-JY, Tovar-Pescador J. Selection of input parameters to model direct solar irradiance by using artificial neural networks. Energy 2005;30:1675–84.
[8] Soares J, Oliveira AP, Božnar MZ, Mlakar P, Escobedo JF, Machado AJ. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique. Appl Energy 2004;79:201–14.
[9] Belaid S, Mellit A. Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. Energy Convers Manage 2016;118:105–18.
[10] Eissa Y, Marpu PR, Gherboudj I, Ghedira H, Ouarda TBMJ, Chiesa M. Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using SEVIRI images. Sol Energy 2013;89:1–16.
[11] Linarez-Rodriguez A, Quesada-Ruiz S, Pozo-Vazquez D, Tovar-Pesacor J. Na evolutionary artificial neural network ensemble model for estimating hourly direct normal irradiances from Meteosat imagery. Energy 2015;91:264–73.
[12] Jian Y. Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. Energy 2009;34:1276–83.
[13] Yadav AK, Chandel SS. Solar radiation prediction using artificial neural network techniques: a review. Renew Sustain Energy Rev 2014;33:772–81.
[14] Raghavendra S, Deka PC. Support vector machine applications in the field of hydrology: a review. Appl Soft Comput 2014;19:372–86.
[15] Azimi R, Ghayekhloo M, Ghofrani M. A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. Energy Convers Manage 2016;118:331–44.
[16] Mohammadi K, Shamshirband S, Anisi MH, Alam KA, Petkovic D. Support vector regression based prediction of global solar radiation on a horizontal surface. Energy Convers Manage 2015;91:433–41.
[17] Alam S, Kaushik SC, Garg SN. Computation of beam solar radiation at normal incidence using artificial neural network. Renew Energy 2006;31:1483–91.
[18] Chen J-L, Li G-S, Wu S-J. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. Energy Convers Manage 2013;75:311–8.
[19] Chen J-L, Li G-S, Xiao B-B, Wen Z-F, Lv M-Q, Chen C-D, et al. Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. Energy Convers Manage 2015;89:318–29.
[20] Tomar RK, Kaushika ND, Kaushik SC. Artificial neural network based computational model for the prediction of direct solar radiation in Indian zone. J Renew Sustain Energy 2012;4:063146. http://dx.doi.org/10.1063/1.4772677.
[21] Kaushina ND, Tomar RK, Kaushik SC. Artificial Neural Network model based on interrelationship of direct, diffuse and global solar radiations. Sol Energy 2014;103:327–42.
[22] Ramli MAM, Twaha S, AL-Turkia YA. Investigating the performance of support vector machine and artificial neural networks in predicting solar radiation on a tilted surface: Saudi Arabia case study. Energy Convers Manage 2015;105:442–52.
[23] Haykin S. Neural networks: a comprehensive foundation. 2nd ed. Hamilton: Prentice Hall; 1998.
[24] Khalil AF, Mckee M, Kemblowski M, Asefa T. Basin scale water management and forecasting using artificial neural networks. J Am Water Resour Assoc 2005;41:195–208.
[25] Fiorin DV, Martins FR, Schuch NJ, Pereira EB. Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. Revista Brasileira de Ensino de Física 2011;33:1309–20 [in Portuguese].
[26] Lorena AC, Jacintho LFO, Siqueira MF, Giovanni R, Lohmann LG, Carvalho ACPLF, et al. Comparing machine learning classifiers in potential distribution modelling. Expert Syst Appl 2011;38:5268–75.
[27] Bishop MC. Neural networks for pattern recognition. Oxford University Press; 1995.
[28] Paniagua-Tineo A, Salcedo-Sanz S, Casanova-Mateo C, Ortiz-García EG, Cony MA, Hernández-Martín E. Prediction of daily maximum temperature using a support vector regression algorithm. Renew Energy 2011;36:3054–60.
[29] Lam JC, Wan KKW, Liu Y. Solar radiation modeling using ANNs for different climates in China. Energy Convers Manage 2008;49:1080–90.
[30] Haykin S. Redes Neurais: Princípios e prática. Trad. Paulo Martins Engel. – 2. Ed. – Porto Alegre. Bookman; 2001 [in portuguese].
[31] Vapnik VN. The nature of statistical learning theory. Springer Science & Business Media; 2013.
[32] Alsina EF, Bortolini M, Gamberi M, Regattieri A. Artificial neural network optimization for monthly average daily global solar radiation prediction. Energy Convers Manage 2016;120:320–9.
[33] Mohammadi K, Shamshirband S, Tong CW, Arif M, Petkovic D, Che S. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers Manage 2015;92:162–71.
[34] Antonanzas-Torres F, Urraca R, Antonanzas J, Fernandez-Ceniceros J, Martinez-de-Pison FJ. Generation of daily global solar irradiation with support vector machines for regression. Energy Convers Manage 2015;96:277–86.
[35] Vapnik VN. Statistical learning theory. New York: Wiley; 1998.
[36] Piri J, Shamshirband S, Dalibor Petkovic D, Tong CW, Rehman MH. Prediction of the solar radiation on the Earth using support vector regression technique. Infrared Phys Technol 2015;68:179–85.
[37] Ramedani Z, Omid M, Keyhani A, Shamshirband S, Khoshnevisan B. Potential of radial basis function based support vector regression for global solar radiation prediction. Renew Sustain Energy Rev 2014;39:1005–11.
[38] Liu J, Zio E. An adaptive online learning approach for support vector regression: online-SVR-FID. Mech Syst Sig Process 2016;76–77:796–809.
[39] Hsu CC, Chang CC, Lin CJ. A practical guide to support vector classificationAvailable from: <http://www.csie.ntu.edu.tw/~cjlin/papers/cuide/guide.pdf>2010.
[40] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011. 2:27:1-27:27.
[41] Hall M, Frank E, Holmes G, Pfahringer B, Reutemrna P, Witten IH. The WEKA data mining software: an update; SIGKDD Explorations, vol. 11, Issue 1; 2009.
[42] Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques, 3rd ed.; 2011 (630p).
[43] Yadav AK, Malik H, Chandel SS. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renew Sustain Energy Rev 2012;31:509–19.
[44] De Souza JL, Lyra GB, Dos Santos CM, Ferreira Junior RF, Tiba C, Lyra GB, et al. Empirical models of daily and monthly global solar irradiation using sunshine

duration for Alagoas State, Northeastern Brazil. Sustain Energy Technol Assess 2016;14:35–45.

[45] Iqbal M. An introduction to solar radiation. New York: Academic Press; 1983.

[46] Cañada J, Pinazo JM, Boscá JV. Determination of Angström's turbidity coefficient at Valencia. Renew Energy 1993;3:621–6.

[47] Escobedo JF, Gomes EN, Oliveira AP, Soares J. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. Renew Energy 2011;36:169–78.

[48] Gueymard CA, Ruiz-Arias JA. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Sol Energy 2015. http://dx.doi.org/10.1016/j.solener.2015.10.010.

[49] Jamieson PD, Porter JR, Wilson DR. A test of the computer simulation model ARC-WHEAT1 on wheat crops grown in New Zealand. Field Crops Res 1991;27:337–50.

[50] Heinemann AB, Van Oor PAJ, Fernandes DS, Maia AHN. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. Bragantia, Campinas 2012;71(4):572–82.

[51] Santos CM, Escobedo JF. Temporal variability of atmospheric turbidity and DNI attenuation in the sugarcane region, Botucatu/SP. Atmos Res 2016;181:312–21.

[52] Chaves M, Escobedo JF. A software to process daily solar radiation data. Renew Energy 2000;19:339–44.

[53] WMO – World Meteorological Organization. Guide to meteorological Instruments and Methods of Observation. WMO-n°8, seventh ed., Geneva, Switzerland; 2008. p. 1–681.

[54] Chen J-L, Liu H-B, Wu W, Xie D-L. Estimation of monthly solar radiation from measured temperatures using support vector machines – a case study. Renew Energy 2011;36:413–20.

[55] Meyer R, Beyer H G, Fanslau J, Geuder N, Hammer A, Hirsch T, et al. Towards standardization of CSP yield assessments. In: SolarPACES, Germany. p. 1–8.

[56] Marquez R, Coimbra CFM. Forecasting of global and direct solar irradiance using stochastic learning methods, group experiments and the NWS data-base. Sol Energy 2011;85:746–56.

[57] Cristianini N, Tylor JS. An introduction to support vector machines and other kernel-based learning methods. UK: Cambridge University Press; 2000.

[58] Han D, Chan L, Zhu N. Flood forecasting using support vector machines. J Hydroinform 2007;9:267–76.

[59] Fernández-Peruchena CM, Blanco M, Bernardos A. Generation of series of high frequency DNI years consistent with RNA annual and monthly long-term averages using measured DNI data. Energy Proc 2014;49:2321–9.

[60] Grantham AP, Pudney PJ, Boland JW, Belusko M. Synthetically interpolated five-minute direct normal irradiance. In: 20th international congress on modelling and simulation, Adelaide, Australia, 1–6 December.

[61] Polo J, Zarzalejo L, Marchante R, Navarro A. A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. Sol Energy 2011;85:1164–70.

[62] Escobedo JF, Gomes EN, Oliveira AP, Soares JR. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. Appl Energy 2009;86:299–309.

[63] Gueymard CA. Temporal variability in direct and global irradiance at various time scales as affected by aerosols. Sol Energy 2012;86:3544–53.