UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

FCA
FACULDADE DE CIÊNCIAS AGRONÔMICAS
UNESP | BOTUCATU

ANGÉLICA CASTILHO PAES

ESTIMATIVA DA IRRADIAÇÃO SOLAR GLOBAL PELOS MODELOS HARGREAVES-SAMANI, BRISTOW- CAMPBELL, REGRESSÕES MÚLTIPLAS E TÉCNICAS MVS E RNA NO MUNICÍPIO DE BOTUCATU - SP

Botucatu

2021

ANGÉLICA CASTILHO PAES

ESTIMATIVA DA IRRADIAÇÃO SOLAR GLOBAL PELOS MODELOS HARGREAVES-SAMANI, BRISTOW- CAMPBELL, REGRESSÕES MÚLTIPLAS E TÉCNICAS MVS E RNA NO MUNICÍPIO DE BOTUCATU - SP

Dissertação apresentada à Faculdade de Ciências Agronômicas da Unesp Câmpus de Botucatu, para obtenção do título de Mestre em Energia na Agricultura.

Orientador(a): Prof. Dr. João Francisco Escobedo

Botucatu

2021

P126e

Paes, Angélica Castilho

Estimativa da irradiação solar global pelos modelos Hargreaves-Samani, Bristow-Campbell, regressões múltiplas e técnicas MVS e RNA no município de Botucatu - SP / Angélica Castilho Paes. -- Botucatu, 2021

75 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agronômicas, Botucatu

Orientador: João Francisco Escobedo

1. radiação solar. 2. modelos estatísticos de estimativa da radiação. 3. regressões múltiplas não lineares. 4. técnicas de aprendizagem de máquinas. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agronômicas, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: ESTIMATIVA DA IRRADIAÇÃO SOLAR GLOBAL PELOS MODELOS DE HARGREAVES-SAMANI, BRISTOW-CAMPBELL, REGRESSÕES MÚLTIPLAS E TÉCNICAS MVS E RNA NO MUNICÍPIO DE BOTUCATU - SP

AUTORA: ANGÉLICA CASTILHO PAES

ORIENTADOR: JOÃO FRANCISCO ESCOBEDO

Aprovada como parte das exigências para obtenção do Título de Mestra em AGRONOMIA (ENERGIA NA AGRICULTURA), pela Comissão Examinadora:

Prof. Dr. JOÃO FRANCISCO ESCOBEDO (Participação Virtual)
Engenharia Rural e Socioeconomia / Faculdade de Ciências Agronômicas de Botucatu - UNESP

Pesquisador Dr. JOSÉ CARLOS FIGUEIREDO (Participação Virtual)
IPMet – Centro de Meteorologia de Bauru / Faculdade de Ciências de Bauru – UNESP

Prof. Dr. SÉRGIO AUGUSTO RODRIGUES (Participação Virtual)
Bioprocessos e Biotecnologia / Faculdade de Ciências Agronômicas de Botucatu - UNESP

Botucatu, 29 de janeiro de 2021

Aos meus amados pais,

Cleide e Manuel, pelo incentivo constante

dedico

AGRADECIMENTOS

Ao Prof. Dr. João Francisco Escobedo, pela oportunidade, pela orientação e pelos ensinamentos.

A Universidade Estadual Paulista, a Faculdade de Ciências Agronômicas e ao Programa de Pós-Graduação em Energia na Agricultura (UNESP/FCA/BOTUCATU) pela oportunidade e estrutura oferecidas para a conclusão de mais uma etapa no meu desenvolvimento profissional.

Aos professores Dr. José Carlos Figueiredo e Dr. Sérgio Augusto Rodrigues por aceitarem compor a banca avaliadora desta dissertação.

Aos colegas de laboratório Maurício, Edson, Adriano e Taiza, pelos ensinamentos, conselhos e contribuições, diretas ou indiretas, para este trabalho.

A Sulei Barros e Katia Albano, membros da equipe gestora da escola CEEJA – Botucatu, e ao estimado amigo Antonio, que possibilitaram a conciliação entre meus deveres profissionais e meus interesses acadêmicos. Muito obrigada!

“Lembre-se de olhar para as estrelas e não para baixo, para os seus pés. Tente achar sentido no que você vê e pergunte sobre o que faz o Universo existir. Seja curioso”.

Sthephen Hawking

RESUMO

No presente trabalho realizou-se um estudo comparativo entre os modelos de estimativas da irradiação solar global (H_G) diária: Hargreaves-Samani (HS); Bristow e Campbell (BC); uma equação de regressão múltipla não linear (RM) desenvolvida neste trabalho e duas técnicas de Aprendizado de Máquina (Máquinas de Vetores de Suporte - MVS e Redes Neurais Artificiais - RNA). A base de dados usada no trabalho foi medida em Botucatu no período dos anos de 1996-2011. Os modelos foram estimados com a base de dados total, sendo que para validação dos modelos foram usados duas bases de dados, denominados ano típico e ano atípico, criados a partir da base total dos dados. O modelo estatístico HS foi estimado pelo método dos mínimos quadrados, utilizado para ajustes lineares, e os modelos BC e RM foram estimados por algoritmos de regressão não linear que utilizam métodos iterativos para a estimativa dos modelos, enquanto que as técnicas, MVS e RNA, foram treinadas com entradas de 4 variáveis meteorológicas totalizando 4 combinações no programa WEKA. Os resultados obtidos na validação dos modelos mostram que: os modelos estatísticos HS, BC podem estimar H_G com coeficientes de correlação da ordem de r = 0,84 , e os valores de rMBE e rRMSE variando de +/- 3,0% e 20,0% - 25,0% respectivamente para anos típicos e atípicos; o modelo RM pode estimar H_G com valor de r da ordem de 0,960, e valores de rMBE e rRMSE variando de +/- 4,0% e 11,0% - 14,0% respectivamente para anos típicos e atípicos; os indicativos estatísticos obtidos para a combinação 4, a de melhor desempenho, pode estimar H_G com valores de r = 0,965 e r = 0,960 e valores para MVS (rMBE = +/- 0,5 %; rRMSE = 12,0 %) e RNA (rMBE = +/- 1,5; rRMSE = 12,0 % para anos típico e atípico). A comparação entre os desempenhos dos modelos HS, BC, RM e as técnicas RNA e MVS na combinação 4, sugere o uso do modelo de RM, quando os dados de RI e UR estejam disponíveis, por ser um método mais prático que o uso das técnicas de aprendizagem de máquinas.

Palavras-chave: Radiação solar. Modelo de Hargreaves-Samani. Modelo de Bristow e Campbell. Técnicas de aprendizado de máquinas.

ABSTRACT

In this paper is described a comparative study between models for estimating daily global solar radiation (H_G): Hargreavis-Samani (HS), Bristow-Campbell (BC), a non-linear multiple regression equation developed in this work and two learning machine techniques (Support Vector Machine - MVS and Artificial Neural Network - RNA). The database used in this work was measured in Botucatu in the period of years of 1996-2011. The models were generated and trained with the total database, whereas for model validation were used two fictitious years, denominated typical and atypical, created from the total database. The statistical models HS was estimated by the least saqures method, used for linear regressions, and the BC and RM models were estimated by algorithms for non linear regression that, use iteratives methods to estimate, MVS and RNA, were trained with input of four meteorological variables totalizing four combinations on the WEKA software. The results show that: the statistical models HS, BC can estimate H_G with coefficient of correlation of the range of r = 0,84 and the values of rMBE and rRMSE varying from +/- 3,0% and 20,0% - 25,0%, respectively, for typical and atypical years; the multivariate equation can estimate H_G with values of r of the order of 0,960, and values of rMBE and rRMSE varying from +/- 4,0% and 11,0% - 14,0%, respectively, for the typical and atypical years; the statistical indicative obtained for combination 4, the with the best performance, can estimate HG with values of r = 0,965 and r = 0,960 and values for MVS (rMBE = +/- 0,5%, rRMSE = 12,0%), and RNA (rMBE = +/- 1,5%; rRMSE = 12,0%, for typical and atypical years). The comparision between the performance of the HS, BC, RM models and RNA and MVS techniques in the conbination 4, suggest the use of the RM model, when the RI and UR datas are availables, for being a method more pratical than the use of the learning machine techniques.

Keywords: Solar radiation. Hargreaves-Samani model. Bristow-Campbell model. Machine Learning Techniques.

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizagem de Máquinas.
ANFIS	Sistema de Inferência Neuro Fuzzy Adaptativo.
GF	Genética Fuzzy
GP	Programação genética.
HS	Hargreaves e Samani.
MLP	Multilayer Perceptron.
MVS	Máquina de Vetor de Suporte.
OMM	Organização Mundial de Meteorologia.
RBF	Função de base radial.
RM	Regressão múltipla não linear.
SMO	Sequential Minimal Optimization.
VCAN	Vórtice Ciclônico em Altos Níveis.
WEKA	Waikato Environment for Knowledge Analysis.

LISTA DE SÍMBOLOS

H_G Irradiação solar global diária (MJ m^{-2} dia^{-1}).

I_G Irradiância solar global (kW m^{-2}).

n Brilho solar (h).

H_0 Irradiação solar no topo da atmosfera (MJ m^{-2} dia^{-1}).

H_d Irradiação solar difusa (MJ m^{-2} dia^{-1}).

H_D Irradiação solar direta (MJ m^{-2} dia^{-1}).

N Fotoperíodo (horas).

RI Razão de insolação (adimensional).

a_{AP}, b_{AP} Coeficientes de calibração do modelo Angström Prescott

a_{HS} Coeficiente de calibração do modelo Hargreaves e Samani.

ΔT_1 Amplitude térmica para o ajuste modelo Hargreaves e Samani (°C).

a_{BC}, b_{BC}, c_{BC} Coeficientes de calibração do modelo Bristow e Campbell.

ΔT_2 Amplitude térmica para o ajuste do modelo Bristow e Campbell.

$T_{m(j)}$ Temperatura mínima do dia j (°C).

$T_{m(j+1)}$ Temperatura mínima do dia subsequente ao dia j (°C)

y Variável dependente da equação de regressão não linear múltipla genérica.

f Função de regressão múltipla que relaciona uma variável dependente a uma ou mais variáveis independente.

x Vetor de variáveis independentes de uma equação regressão não linear múltipla genérica.

θ Coeficientes de uma equação de regressão não linear múltipla genérica.

ε Erro aleatório de uma equação de regressão não linear múltipla genérica.

UR Umidade relativa (%).

H_0	Irradiação solar no topo da atmosfera (MJ m^{-2} dia^{-1}).
I_0	Constante solar (W m^{-2}).
DJ	Dia juliano.
φ	Latitude local (graus).
δ	Declinação solar (graus).
ω_s	Ângulo horário (graus).
a_{SO}, b_{SO}, c_{SO}	Coeficientes da equação de regressão não linear múltipla proposta por Swartman e Ogunlade.
$a_{RM}, b_{RM}, c_{RM}, d_{RM}, e_{RM}$	Coeficiente de calibração do modelo RM.
C, γ	Parâmetros de ajuste da Máquina de Vetor de Suporte.
rRMSE	Raiz quadrada do erro médio relativa (%).
H_m	Valores medidos.
H_e	Valores estimados.
$\bar{H}_m$	Valor médio medido.
m	Número de observação de dados.
x_i	Variáveis inseridas na camada de entrada da RNA.
w_{ij}	Pesos sinápticos entre a camada de entrada e a camada oculta da RNA.
w_0	Limiar de ativação da RNA.
u	Sinal transmitido pela camada de saída da RNA.
u_j	Sinal transmitido para a camada oculta da RNA.
ϕ	Função de ativação da RNA.
μ	Somatória das variáveis de entrada multiplicadas pelos pesos sinápticos.
L	Taxa e aprendizagem da RNA.
M	Momentum da RNA.
MBE	Tendência do erro médio (MJ m^{-2} dia^{-1}).
rMBE	Tendência do erro médio relativa (%).

RMSE Raiz quadrada do erro médio (MJ m^{-2} dia^{-1}).

r Coeficiente de correlação (adimensional).

$\overline{H}_e$ Valor médio estimado.

R^2 Coeficiente de determinação (adimensional).

d Índice de concordância de Willmot (adimensional).

SUMÁRIO

1 INTRODUÇÃO..... 23
2 REVISÃO DE LITERATURA..... 26
2.1 Radiação solar..... 26
2.2 Modelos estatísticos..... 27
2.3 Modelos de regressão múltipla..... 29
2.4 Aprendizagem de máquinas..... 31
2.4.1 Rede neural artificial..... 31
2.4.2 Máquina de vetor de suporte..... 32
3 MATERIAIS E MÉTODOS 34
3.1 Local e clima..... 34
3.2 Instrumentos e medidas..... 35
3.3 Modelos estatísticos para a estimativa de H_G..... 36
3.3.1 Hargreaves e Samani..... 37
3.3.2 Bristow e Campbell..... 37
3.4 Regressões múltiplas..... 37
3.5 Aprendizado de máquinas..... 38
3.5.1 Rede neural artificial..... 39
3.5.2 Máquina de vetor de suporte..... 42
3.6 Validação dos modelos..... 43
4 RESULTADOS E DISCUSSÕES..... 46
4.1 Modelos estatísticos..... 46
4.1.1 Obtenção e validação da equação HS..... 46
4.1.2 Obtenção e validação da equação BC..... 50
4.2 Regressão múltipla não linear..... 53
4.3 Aprendizado de máquinas..... 55
4.3.1 Treinamento validação das técnicas de AM..... 55
4.3.2 Comparação do desempenho das técnicas de AM com resultados da literatura..... 60
4.3.3 Comparação do desempenho das técnicas de AM com os modelos estatísticos..... 63
4.3.4 Comparação do desempenho das técnicas de AM com a

equação RM..... 64
5 CONCLUSÃO..... 65
REFERÊNCIAS..... 67

1 INTRODUÇÃO

A irradiação solar global (H_G), principal fonte de energia do planeta, é responsável por diversos processos físicos, biológicos e bioquímicos, participando do balanço de energia, ciclo hidrológico, e afetando diversos parâmetros meteorológicos (PIRI; KISI, 2014). O conhecimento sobre H_G tem importância fundamental no desenvolvimento de sistemas de energia fotovoltaica e estudos na área da agricultura (MEHDIZADEH, 2016).

Em relação ao potencial fotovoltaico, o Brasil apresenta médias anuais de irradiância solar global (I_G) entre 4,53 $kWh.m^{-2}$ (região Sul) e 5,49 $kWh.m^{-2}$ (região Nordeste). Estes valores são maiores do que as médias anuais para países como Espanha, Itália, Alemanha, Portugal e França, onde a tecnologia fotovoltaica está bem estabelecida. Os valores das médias mensais de I_G encontrados na região Nordeste superam os valores encontrados nos países ibéricos, e possuem uma variabilidade menor. A região Sul possui um comportamento similar aos países europeus, em relação ao valor anual e a variabilidade mensal dos valores, devido ao fato de se encontrar em latitudes maiores (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS, 2017).

O Brasil, país emergente, de dimensões continentais e com diferentes regiões climáticas, possui poucas estações meteorológicas mecânicas e automáticas distribuídas em todo território nacional, medindo continuamente H_G. Por essa razão existem poucas informações de mapas solarimétricos de H_G disponíveis na área de especialização. Para suprir a falta de conhecimento de H_G por meio de medidas, muitos pesquisadores no Brasil e nos demais países com a mesma necessidade, vem desenvolvendo modelos que permitem estimar H_G através de variáveis mais comumente encontradas nas estações meteorológicas. Esses modelos podem ser classificados como de regressão (lineares ou não lineares), físicos e técnicas de Aprendizado de Máquina (AM) (HARGREAVES; SAMANI, 1982; WINSLOW et al., 2001; SHARIFI et al., 2016; LANDERAS et al., 2012, CHEN et al., 2013).

Os modelos estatísticos realizam a estimativa de H_G relacionando esta variável com variáveis meteorológicas locais, restringindo a solução para a região estudada ou para locais de clima semelhante. Os modelos físicos (transferência radiativa) exigem medidas com precisão de variáveis meteorológicas na superfície e na atmosfera, necessitando de dados sobre a quantidade de nuvens e a concentração de aerossóis e outros gases atmosféricos, o que torna estes modelos complexos e menos acessíveis do que os estatísticos (MARTINS et al., 2004). Recentemente as técnicas de AM vêm sendo testadas para a estimativa de H_G a partir de diversas variáveis meteorológicas, apresentando resultados superiores aos modelos estatísticos (KUMAR et al., 2015).

O principal modelo estatístico linear é o de Angstrom-Prescott (1924), o qual estima H_G a partir do brilho solar (n), normalmente monitorado nos postos meteorológicos. Para locais ou estações secundárias onde n não é monitorado, são utilizados os modelos que estimam H_G a partir de dados de temperatura do ar, como os modelos de Hargreaves e Samani (1982) e Bristow e Campbell (1984).

Outros pesquisadores desenvolveram seus modelos ou modificaram modelos já existentes e testaram para diversas localidades, como o modelo desenvolvido por Donatelli e Campbell (1998), que trouxe modificações ao modelo de Bristow e Campbell (1984), introduzindo uma variável para correção de efeitos de sazonalidade; Winslow et al. (2001) também propôs modificações no modelo de Bristow e Campbell (1984), introduzindo as variáveis pressão de vapor de saturação para temperatura mínima e máxima; Annandale et al. (2000) propôs uma modificação no modelo de Hargreaves e Samani (1982), incluindo a altitude local na equação. Outros trabalhos como os desenvolvidos por Ertekin e Yaldiz (1999) e Falayi et al. (2008) testaram regressões múltiplas, utilizando diversas variáveis meteorológicas como: umidade relativa, razão de insolação, temperatura média, entre outras, com resultados similares ou melhores do que os modelos lineares.

As técnicas de Aprendizagem de Máquina (AM) vêm sendo usadas na geração de modelos de estimativa de H_G em função das medidas de diversas variáveis meteorológicas, com bons resultados como mostra os estudos de Elizondo et al.

(1994), Bocco et al. (2006); Linarez-Rodríguez et al. (2011), Chen et al. (2013), Chen et al. (2015), Mohammadi et al. (2016), Wan et al. (2008), entre outros. As técnicas de AM são algoritmos computacionais capazes de aprender a partir de dados oferecidos, classificando estes dados e generalizando a solução para classificações futuras, sendo que quanto maior o banco de dados e mais precisos os dados fornecidos, melhor será o aprendizado (MOHRI et al., 2018; KOTSIANTIS, 2007).

Das diversas técnicas disponíveis como rede neural artificial (RNA), programação genética (GP), máquina de vetor de suporte (MVS), sistema de inferência neuro-fuzzy adaptativo (ANFIS), entre outras, a mais utilizada para gerar estimativas de H_G é a RNA, porém estudos apontam que a MVS tem apresentado resultados superiores (MOHAMMADI et al., 2015; AYODELE et al., 2016; CHEN et al., 2013).

Existem poucos estudos comparativos entre modelos AM e modelos estatístico lineares ou não lineares para estimativa de H_G, usando a temperatura como variável meteorológica de entrada nos modelos. Essas pesquisas são de fundamental importância para a maioria das regiões do país onde a radiação solar global não é monitorada, e sim a temperatura. Diante dessa necessidade, o presente trabalho teve o objetivo geral de realizar um estudo comparativo entre diferentes modelos de estimativa da irradiação solar global (H_G) a partir de dados meteorológicos do município de Botucatu. Os objetivos específicos deste trabalho foram: estudar o desempenho dos modelos estatísticos de Hargreaves - Samani (HS) e Bristow - Campbell (BC); estudar o desempenho de duas técnicas de Aprendizagem de Máquinas (AM): Máquina Vetor de Suporte (MVS) e Redes Neurais Artificiais (RNA) para quatro diferentes arquiteturas de variáveis meteorológicas; comparar os resultados das técnicas de AM com os modelos estatísticos; propor e estudar o desempenho de um modelo de regressão não linear múltiplo (RM) e comparar os resultados deste novo modelo com as técnicas de AM.

2 REVISÃO DE LITERATURA

2.1 Radiação Solar

A irradiação solar é a principal fonte de energia para os processos físicos, biológicos e bioquímicos que ocorrem no planeta, responsável pelo ciclo hidrológico e pelo balanço de energia, principal variável meteorológica, permitindo a circulação atmosférica e trocas de energia entre a Terra e o universo (PIRI; KISI, 2014; ALLEN, 1997; MOOJEN et al.; 2012).

Aproximadamente 95% da radiação emitida pelo Sol tem comprimento de onda entre 0,3 μm e 2,4 μm, considerada uma irradiação de onda curta (MACAGNAN, 2010), sendo chamada de irradiação solar no topo da atmosfera (H_0, MJ m^{-2} dia^{-1}). Ao atingir a atmosfera H_0 sofre os fenômenos de reflexão, absorção e espalhamento. Parte da irradiação que atinge a atmosfera reflete para o espaço; o que não é refletido entra na atmosfera, podendo ser absorvida, aquecendo o ar, espalhada, atingindo a superfície de forma indireta, ou transmitida diretamente para a superfície (RONDÓN et al., 2020).

Os elementos constituintes da atmosfera absorvem a energia respeitando certo intervalo no espectro eletromagnético, sendo que o gás oxigênio e o ozônio absorvem intensamente a radiação ultravioleta, o vapor de água e o dióxido de carbono absorvem a região do infravermelho, enquanto que a região da luz visível do espectro eletromagnético sofre pouca atenuação decorrente da absorção, sendo chamada de “janela atmosférica” da radiação solar (PALÁCIOS; 2017). Outros constituintes da atmosfera, como o gás metano e o óxido nitroso, também contribuem para a absorção. Este fenômeno ocorre em vários níveis da atmosfera, com a absorção do ultravioleta ocorrendo na mesosfera, o infravermelho sendo absorvido na baixa estratosfera e troposfera, e as frequências de irradiação intermediárias sendo absorvidas na estratosfera, pelo ozônio (ECHER, et al.; 2001).

O espalhamento da irradiação solar é um fenômeno que depende da razão entre a dimensão da partícula e o comprimento de onda da irradiação. No caso das partículas muito menores do que o comprimento de onda da irradiação, que podem ser moléculas ou partículas muito pequenas, o fenômeno é chamado de espalhamento Rayleigh, em que ocorre perfeita simetria entre a irradiação

espalhada e o plano normal ao feixe incidente. O espalhamento Rayleigh depende inversamente da quarta potência do comprimento de onda, tornando-se desprezível para irradiações de comprimento de onda maior do que 0,7 μm (PALÁCIOS, 2017). Nos casos em que a partícula tem dimensões da mesma ordem de grandeza que o comprimento de onda da irradiação, conhecidas como aerossóis, o fenômeno é chamado de espalhamento Mie, sendo mais difícil de ser modelado por depender da forma e constituição dos aerossóis, sendo que estas partículas apresentam grande variabilidade. Este fenômeno não depende do comprimento de onda da irradiação, porém é influenciado por diversas outras variáveis, tornando-se um fenômeno difícil de avaliar experimentalmente. Os métodos para lidar com o espalhamento Mie consistem em estimativas dos coeficientes exigidos para a resolução do problema ou medidas locais, portanto devem ser tratados com aproximações (ALMEIDA; 1998).

Esta parcela da irradiação que sofre processos de espalhamento atinge a superfície de forma indireta, sendo chamada de radiação solar difusa (H_d, MJ m^{-2} dia^{-1}). Aproximadamente 25 % da irradiação solar que atinge a Terra é transmitida diretamente para a superfície, chamada de irradiação solar direta (H_D, MJ m^{-2} dia^{-1}). A soma das parcelas difusa e direta é chamada de irradiação solar global (H_G, MJ m^{-2} dia^{-1}) (CROTTI; RAMPINELLI, 2018).

2.2 Modelos estatísticos

Devido às dificuldades de se medir H_G, os pesquisadores desenvolveram modelos estatísticos para estimar H_G por meio de variáveis meteorológicas. Estes modelos estatísticos podem ser agrupados em quatro categorias: modelos de brilho solar, modelos de nebulosidade, modelos de temperatura e modelos baseados em outras variáveis meteorológicas (BESHARAT; 2013). O primeiro modelo estatístico e o mais encontrado em trabalhos da área foi desenvolvido por Angström e Prescott (1924) (AP), e realiza a estimativa de H_G a partir de medidas de n, e valores calculados de H_0 e fotoperíodo (Eq. 1):

$$\frac{H_G}{H_0} = a_{AP} + b_{AP}\left(\frac{n}{N}\right) \quad (1)$$

onde $\frac{H_G}{H_0}$ é chamado de transmissividade atmosférica, $\frac{n}{N}$ é a razão de insolação, a_{AP} e b_{AP} são os coeficientes que devem ser estimados para o local de estudo, sendo que a_{AP} representa a transmissividade mínima, e o coeficiente b_{AP}, quando somado a a_{AP} representa a transmissividade atmosférica de céu aberto (máxima) (SILVA; 2016).

Porém os modelos baseados em RI e em nebulosidade dependem de dados de brilho solar, variável menos comum de se encontrar nas estações meteorológicas do que a temperatura do ar. Portanto é interessante desenvolver e validar modelos que estimam H_G a partir de temperatura do ar, visto que a escolha de uma modelo de estimativa para certa região deve ser feita levando em conta que variáveis estão disponíveis no local (BESHARAT; 2013; LI et al.; 2013).

Os modelos desenvolvidos a partir de dados de temperatura consideram que a variável que mais influencia os valores de temperatura é H_G. A irradiação que atinge a superfície é absorvida, causando a evaporação da água (calor latente) e o aquecimento da superfície e do ar (calor sensível), sendo possível utilizar os valores de amplitude térmica diária (ΔT) para estimar H_G (LANDERAS et al; 2012; LÊDO et al.; 2012), com ΔT exercendo, indiretamente, o papel de indicativo da nebulosidade, principal variável na determinação da quantidade de irradiação que atinge a superfície (ALLEN; 1997; LÊDO et al.; 2012).

O primeiro modelo baseado em dados de ΔT foi desenvolvido por Hargreaves e Samani (1982) (HS), capaz de estimar H_G a partir de dados de H_0 e da amplitude térmica (ΔT_1), Eq. (2).

$$H_G = H_0 \,.\, a_{HS} \,.\, (\Delta T_1)^{0,5} \tag{2}$$

onde a_{HS} é o coeficiente de calibração do modelo, e ΔT_1 é a amplitude térmica para a aplicação do modelo HS (Eq. 3).

$$\Delta T_1 = T_M - T_m \tag{3}$$

O modelo de Bristow e Campbell (1984) (BC) apresenta uma relação exponencial entre H_G e ΔT (Eq. 4):

$$H_G = H_0 \,.\, a_{BC} \left[1 - exp\left((-b_{BC}) \,.\, \Delta T_2^{c_{BC}}\right)\right] \tag{4}$$

sendo ΔT_2 a amplitude térmica adotada no modelo BC calculada pela Eq. (5), que realiza uma média entre a temperatura mínima do dia em questão ($T_{m(j)}$) e a temperatura mínima do dia seguinte ($T_{m(j+1)}$).

$$\Delta T_2 = T_M - \left(\frac{T_{m(j)} + T_{m(j+1)}}{2}\right) \quad (5)$$

O coeficiente a_{BC} representa H_G máxima em dia de céu aberto, b_{BC} e c_{BC} controlam a relação entre a_{BC} e ΔT_2, ou seja, determina o quanto H_G varia em resposta a variação de ΔT (CHEN et al., 2004).

Winslow et al. (2001) desenvolveu um modelo que utiliza indiretamente valores de temperatura, através de dados de pressão de vapor de saturação a temperatura máxima e pressão de vapor de saturação à temperatura mínima, incluiu um fator de correção para o efeito dos aerossóis, do ozônio e do vapor de água atmosféricos na transmitância atmosférica, além de incluir outro fator de correção para as diferenças na duração dos dias.

Outros pesquisadores realizaram adaptações nos modelos citados, como Annandale et al. (2002), que adaptou o modelo HS, incluindo a altitude local na equação; Almorox et al. (2011) adicionou ao modelo HS valores de pressão de vapor de saturação a temperatura máxima e mínima; Hunt et al (1998) modificou o modelo HS e propôs a inclusão dos dados de precipitação diária; Donatelli e Campbell (1998) modificaram o modelo BC, incluindo um fator de correção de sazonalidade para locais em latitude médias; a modificação proposta por Goodin (1999) foi a divisão de ΔT_2 por H_0, com o objetivo de acomodar uma maior variação de H_G; Meza e Varas (2000) fixaram os valores dos coeficientes a_{BC} e c_{BC} do modelo BC, adotando a_{BC} =0,75 e c_{BC}=2 estimando apenas o coeficiente b_{BC}.

2.3 Modelos de regressão múltipla não linear

Além dos trabalhos citados acima, alguns autores desenvolveram modelos de estimativa a partir das variáveis disponíveis para o local de estudo, utilizando métodos de regressão múltipla. Regressão múltipla é um conjunto de métodos estatísticos que permite estudar a relação entre uma variável dependente e uma ou

mais variáveis independentes, e pode ser utilizado em diversas áreas de estudo, como ciências sociais, biologia e física (BERRY; FELDMAN, 1985).

Este método permite conhecer a influência das variáveis independentes sobre o comportamento da variável dependente, escolher as variáveis independentes que mais influenciam a variável dependente, predizer os valores da variável dependente e avaliar as incertezas e desvios presentes no processo (TEIXEIRA Jr. et al.; 2011). Os modelos de regressões múltiplos podem ser representados na forma de Eq. (6) encontrada em Mattos (2013):

$$y = f(x, \theta) + \varepsilon \quad (6)$$

onde o vetor x é composto pelas as variáveis independentes, θ são os parâmetros a serem estimados, f é a função que determina a relação entre as variáveis independentes (x) e a variável dependente (y), e ε são os erros aleatórios. Quando se observa uma relação não linear entre os parâmetros, este modelo é chamado de modelo de regressão não linear múltiplo (YASAR et al.; 2012).

Diversos autores utilizaram técnicas de regressão múltipla, desenvolvendo modelos lineares e não lineares, e combinando diversas variáveis meteorológicas. Falayi et al. (2008) desenvolveu um modelo para estimar H_G mensal para Onne (Nigéria), utilizando dados de RI, temperatura e umidade relativa (UR), adaptando o modelo AP com a inclusão da temperatura e a umidade, adaptação que gerou aumento na precisão das estimativas. Almorox et al. (2011) desenvolveu modelos de regressão não linear para Aranjuez (Espanha) a partir de dados de H_0, pressão de vapor de saturação à temperatura mínima e máxima, dados de precipitação transformados e UR, e comparou com modelos frequentemente utilizados, como os modelos HS e BC, e concluiu que modelos baseados em dados de temperatura que incluem dados de precipitação apresentam resultados mais precisos.

Bilgili e Ozgoren (2011) desenvolveram modelos de regressão múltiplos, lineares e não lineares, utilizando dados de brilho solar, temperatura atmosférica, velocidade do vento, mês do ano e dia do mês, sendo que para ambas as situações os modelos de melhor desempenho foram aqueles que utilizavam todas as variáveis disponíveis, com os modelos não lineares apresentando melhores resultados do que os lineares. O modelo de regressão múltiplo não linear desenvolvido por Maghrabi (2009) para

estimar H_G para Tabouk (Arábia Saudita) utilizou dados de temperatura do ar, umidade relativa, temperatura de ponto de orvalho, pressão de vapor, cobertura do céu e vapor de água precipitável, e apresentou melhor desempenho comparado com modelos encontrados na literatura como os modelos HS e BC.

2.4 Aprendizagem de máquinas

A aprendizagem de máquinas (AM) é uma área de estudo da inteligência artificial, que combina conceitos de ciência da computação, probabilidade, estatística e otimização com o objetivo de desenvolver métodos computacionais capazes de fazer previsões e melhorar o desempenho a partir da experiência adquirida pela interação com o ambiente, ou seja, por meio de dados referentes ao problema fornecidos ao sistema (MOHRI et al., 2018). Estes sistemas devem ter a habilidade de se adaptar a mudanças externas, e predizer o futuro a partir de informações fragmentadas e incertezas (BONACCORSO, 2017).

Na área da agricultura as técnicas de AM são utilizadas em diversas atividades como: detecção de doenças, detecção de plantas daninhas, manejo da irrigação, manejo do solo, entre outras (LIAKOS et al.; 2018). Para a estimativa de H_G várias técnicas de AM foram testadas como a rede neural artificial (RNA), amplamente testada para diversos locais do mundo; a máquina de vetor de suporte (MVS), que tende a apresentar resultados melhores do que a RNA; o Sistema de Inferência Adaptativo Neuro – Fuzzy (ANFIS); Genética Fuzzy (GF), entre outras (ANTONANZAS-TORRES et al., 2015).

2.4.1 Rede Neural Artificial

As redes neurais artificiais (RNA) são consideradas uma subárea das técnicas de aprendizagem de máquinas, e seu desenvolvimento se iniciou com as tentativas de compreender o funcionamento do cérebro. Em 1943 McCulloch e Pitts desenvolveram um modelo matemático binário para simular a atividade neural, em que o neurônio recebia sinais excitatórios e inibitórios, e emitia um sinal de saída quando sinal excitatório fosse maior do que o inibitório (SONDAK; SONDAK, 1989). Em 1946, Hebb apresentou a ideia de modificação sináptica, afirmando que o

conhecimento era armazenado nas conexões dos neurônios, e que o processo de aprendizagem modifica as conexões e altera os efeitos excitatórios e inibitórios das entradas (GARSON, 1998, SONDAK; SONDAK, 1989).

Em 1958 Rosemblatt criou um modelo computacional de classificação linear, para analisar o funcionamento de uma rede neural, com elementos de processamento conectados, chamados perceptrons. Os trabalhos de Marvin Minsky e Seymour Papert demonstraram que o modelo de Rosenblatt apresentava utilidade limitada por não ser capaz de lidar com diversos problemas práticos que exigiam soluções não lineares (KULIK, 2015; GARSON, 1998). Em 1982, Hopfield introduziu funções não lineares ao sistema, tornando as redes neurais capazes de resolver problemas não lineares, o que ampliou as possibilidades de aplicação da RNA (GARSON, 1998).

As aplicações desta técnica são diversas, podendo ser utilizadas para predição, aproximação de funções ou classificação de padrões (GARDNER; DORLING, 1998). Em estudos ambientais a RNA tem diversas aplicações para estimativas, como, por exemplo: a estimativa da evapotranspiração de referência (ZANETTI et al., 2007); previsões da disponibilidade de recursos eólicos para certa localidade (TEIXEIRA Jr. Et al.2011); levantamentos sobre a disponibilidade de energia solar em dada região (FIORIN et al., 2011); previsão de queimadas e incêndios (VIGANÓ et al., 2017); entre outros diversos estudos.

2.4.2 Máquina de Vetor de Suporte

A MVS é uma técnica de aprendizagem de máquinas baseada na teoria de aprendizagem estatística e no princípio da minimização do risco estrutural, ideias desenvolvidas por Vapnik e Chervonenkis (1964) (BELAID; MELLIT, 2016). Esta técnica pode ser aplicada em problemas de classificação, regressão e previsão (CHEN; LI, 2014). Devido ao fato de lidar com problemas de classificação não lineares com generalização satisfatória e evitando super-ajustes, a técnica MVS conquistou popularidade (TORRES et al., 2015).

A técnica MVS tem sido utilizada em diversas áreas do conhecimento como em computação, estudos em hidrologia e ambiente (MOHAMMADI et al., 2015); esta técnica também é aplicada em estudos ambientais, como em previsões climáticas

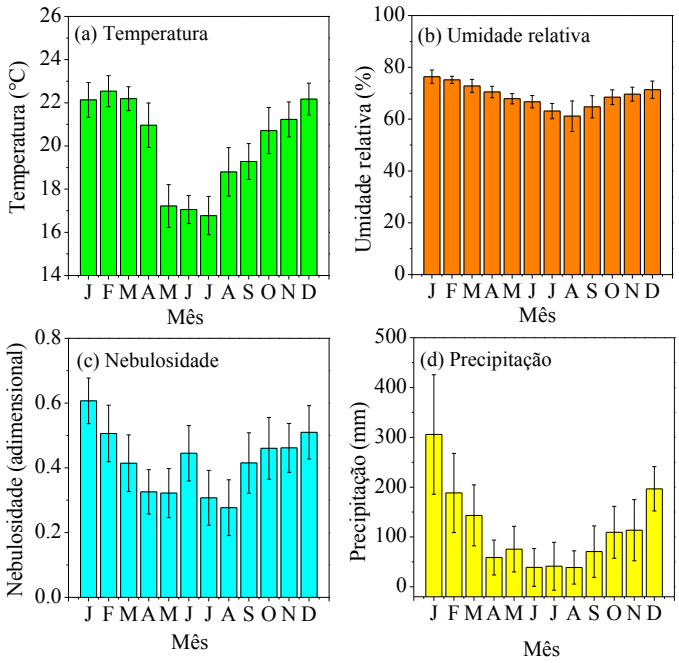
(RADHIKA; SHASHI, 2009), previsão de tempestades (RAJASEKARAN et al., 2008), estimativa da concentração de ozônio (FENG et al., 2011).

3 MATERIAIS E MÉTODOS

3.1 Local e clima

A cidade de Botucatu está localizada no Estado de São Paulo, a 22°53’ de latitude Sul e 48°26’ de longitude Oeste. A altitude varia de 650 m a 950 m em relação ao nível do mar, e o clima de Botucatu é classificado como Cwa, pelo método de Köppen, sendo um clima temperado quente (mesotérmico) (RAMOS et al., 2011). A Figura 1 (a, b, c, d) apresenta as séries climáticas da temperatura, umidade relativa, precipitação e nebulosidade para Botucatu, no período entre 1970 – 2008.

Figura 1 – Médias mensais da temperatura (a), umidade relativa (b), nebulosidade (c) e precipitação (d) medidas no período de 1970 – 2008



Fonte: adaptado de Silva (2016, p. 14)

A temperatura média mensal varia de 16,8 °C em julho a 22,5 °C em fevereiro. A umidade relativa tem seu maior valor em janeiro (76,4 %), e o menor valor em agosto (61,1 %). A nebulosidade representa a fração de horas de céu encoberto por dia, sendo o maior resultado encontrado para a nebulosidade foi para

o mês de janeiro (0,61), e o menor para o mês de agosto (0,28). A precipitação tem o maior valor em janeiro com 305,7 mm, e o menor valor para agosto, com 38,8 mm.

Os dados de umidade, nebulosidade e precipitação refletem as características climáticas da região que apresenta duas estações bem definidas, com verão quente e chuvoso e inverno seco e frio. Entre 75 e 90% das chuvas ocorrem entre outubro e março, e entre abril e setembro ocorre o período seco, apresentando níveis pluviométricos abaixo de 100 mm (TERAMOTO; ESCOBEDO, 2012). O período chuvoso que ocorre na região está relacionado com a presença do Vórtice Ciclônico em Altos Níveis, no verão brasileiro. Esse sistema se desloca no oceano Atlântico, entre a África e o Nordeste do Brasil, atingindo o continente, com uma extensa banda de nebulosidade, associada a chuvas, na parte oeste do VCAN (KOUSKY, 1981). Os sistemas frontais, que são os encontros de duas massas de ar de temperatura e densidade diferentes, atuam na região durante todo o ano, permanecendo no máximo dois dias (ALVARENGA, 2012).

3.2 Instrumentos e variáveis analisadas

Os dados meteorológicos utilizados neste trabalho foram obtidos no período de 1996-2011, na estação de radiometria solar, instalada na Faculdade de Ciências Agronômicas (FCA), Universidade Estadual Paulista (UNESP). As variáveis utilizadas neste trabalho foram: irradiância solar global, (I_G, W m^{-2}); brilho solar, definido como a quantidade de horas de brilho solar sem a presença de nuvens (n, horas); temperatura máxima (T_M, °C); temperatura mínima (T_m, °C) e umidade relativa (UR, %). As medidas de I_G foram realizadas por um piranômetro Eppley PSP; n foi medido por um heliógrafo Campbell – Stokes; T_M e T_m foram medidas por um termômetro de mercúrio e um termômetro de álcool, respectivamente; e UR foi medida por um higrômetro, seguindo as recomendações da Organização Mundial de Meteorologia (OMM).

A aquisição de medidas de H_G foi realizada com o auxílio de um datalogger CR23X da Campbell Scentific, com frequência de operação de 1 Hz, armazenando o valor médio a cada cinco minutos. As medidas de irradiância (W m^{-2}) foram integradas por meio de programas para a análise de irradiação (CHAVES;

ESCOBEDO, 2000), para a obtenção do valor da irradiação solar global diária (H_G, MJ m^{-2} dia^{-1}).

A irradiação solar no topo da atmosfera (H_0, MJ m^{-2} dia^{-1}) foi calculada a partir das equações encontradas em Santos (2016), onde I_0 é a constante solar (1367 W m^{-2}), DJ é o dia juliano (sendo que para o dia primeiro de janeiro, DJ = 1, e para o dia 31 de dezembro, DJ = 365), φ é a latitude local (graus), δ é a declinação solar (graus) (Eq. 8), e ω_s é o ângulo horário (graus) (Eq. 9).

$$H_0 = I_0 \,.\left[1 + 0{,}033 \,.\cos\frac{2\pi DJ}{365}\right].[(sen\varphi \,.sen\delta) + (cos\varphi \,.cos\delta \,.sen\omega_s)] \quad (7)$$

$$\delta = 23{,}45 \,.sen\left[\frac{360\,.\,(DJ+284)}{365}\right] \quad (8)$$

$$\omega_s = cos^{-1}(-tg\varphi \,.\, tg\delta) \quad (9)$$

O fotoperíodo é a quantidade máxima possível de horas de brilho solar (N, horas), calculado a partir da Eq. (10):

$$N = \frac{2\,.\,\omega_s}{15} \quad (10)$$

A razão de insolação é a relação entre a quantidade real (n) e a quantidade máxima de horas de brilho solar (N) (CHAVES et al., 1999), apresentado na Eq. (11):

$$RI = \frac{n}{N} \quad (11)$$

3.3 Modelos estatísticos para a estimativa de H_G

Os dados utilizados no desenvolvimento dos modelos, tanto estatísticos (modelos de regressão clássicos para a estimativa de H_G e modelo de regressão não linear múltiplo), quanto de aprendizado de máquinas (RNA e MVS), consistem em medidas diárias das variáveis meteorológicas coletadas entre os anos 1996 – 2011 em Botucatu – SP.

A estimativa dos modelos clássicos, bem como a plotagem dos gráficos apresentados neste trabalho e o cálculo dos indicativos estatísticos para a validação dos modelos foram realizados por meio do software Microcal™ Origin 6.0 ®.

3.3.1 Hargreaves e Samani

O modelo HS (HARGREAVES; SAMANI, 1985) propõe relacionar H_G com ΔT_1, onde a_{HS} é o coeficiente de calibração a ser estimado para Botucatu. O cálculo de $\sqrt{\Delta T_1}$ realizado previamente possibilitou a estimativa do coeficiente a_{HS} a partir de uma regressão linear, pelo método dos mínimos quadrados (MICROCAL SOFTWARE, 1999).

3.3.2 Bristow e Campbell

O modelo BC (BRISTOW; CAMPELL, 1984) propõe relacionar H_G com ΔT_2, onde a_{BC}, b_{BC} e c_{BC} são os coeficientes de calibração a serem estimados para Botucatu. O algoritmo disponível para modelos não lineares é o Levenberg-Marquart, que oferece um método de estimativa iterativo, em que valores iniciais dos coeficientes devem ser fornecidos para que estimativa seja realizada. A determinação adequada destes valores torna a estimativa do modelo mais rápida e confiável (MICROCAL SOFTWARE, 1999). Os valores iniciais escolhidos foram encontrados no trabalho de Quej et al. (2016).

3.4 Regressões múltiplas

Para desenvolver o modelo de regressão não linear múltiplo, proposto neste trabalho, foram utilizados como referência modelos estatísticos encontrados na literatura, para os quais a ordem de grandeza dos parâmetros já é conhecida (ANGSTRÖN, 1924; SWARTMAN; OGUNLADE, 1967; HARGREAVES, 1985). O modelo HS foi utilizado como referência para a inclusão dos valores de ΔT_1 no novo

modelo. O modelo de Angström - Prescott foi utilizado como referência para incluir a variável RI. Para a inclusão de UR foi utilizado um modelo que relaciona H_G com RI e UR, encontrada em Swartman e Ogunlade (1967) e apresentada na Eq. (12).

$$H_G = a_{SO} \,.\, RI^{bso} . UR^{cso} \tag{12}$$

Assim, a Eq. (13) apresenta o modelo de regressão não linear desenvolvido neste trabalho e validado para Botucatu.

$$H_G = a_{RM} . H_0 \,.\, (\Delta T^{b_{RM}}) + c_{RM} \,.\, H_0 \,.\, (RI^{d_{RM}}) \,.\, (UR^{e_{RM}}) + \varepsilon \tag{13}$$

onde $a_{RM}, b_{RM}, c_{RM}, d_{RM}$ e e_{RM} são os coeficientes a serem estimados.

A estimativa deste modelo foi realizada no R, um ambiente com ferramentas para desenvolver pesquisas em estatística, como, por exemplo, regressões lineares e não lineares, classificações, testes estatísticos, etc. Este ambiente é um software livre e de código aberto (R CORE TEAM, 2018). O algoritmo disponível para a estimativa dos modelos não lineares é o Gauss-Newton, que também utiliza um método iterativo para a estimativa dos coeficientes do modelo. Portanto, assim como na estimativa dos coeficientes do modelo BC, neste caso é preciso fornecer valores iniciais, para que o algoritmo realize o processo iterativo na estimação dos parâmetros.

3.5 Aprendizagem de máquinas

O software utilizado para realizar o aprendizado de máquinas foi o WEKA (Waikato Environment for Knowledge Analysis), desenvolvido pela Universidade de Waikato, Nova Zelândia, e possui diversos procedimentos para aplicação de: regressão linear, árvore de decisão, redes neurais artificiais, máquina de vetor de suporte, entre outras.

Para as técnicas de aprendizado de máquinas foram testadas quatro combinações das variáveis de entrada (Tabela 1) para avaliar que combinação de variáveis apresentava os melhores resultados.

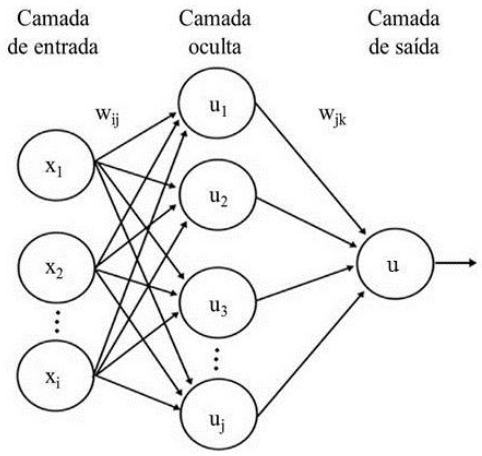
Tabela 1 – Variáveis utilizadas nas combinações testadas para as técnicas RNA e MVS

Combinações	Variáveis utilizadas
RNA1	H_0, $(\Delta T_1)^{0,5}$ e H_G
MVS1	H_0, $(\Delta T_1)^{0,5}$ e H_G
RNA2	H_0, $(\Delta T_1)^{0,5}$, RI e H_G
MVS2	**H_0, $(\Delta T_1)^{0,5}$, RI e H_G**
RNA3	H_0, $(\Delta T_1)^{0,5}$, UR e H_G
MVS3	H_0, $(\Delta T_1)^{0,5}$, RI, UR e H_G
RNA4	H_0, $(\Delta T_1)^{0,5}$, RI, UR e H_G
MVS4	H_0, $(\Delta T_1)^{0,5}$, RI, UR e H_G

3.5.1 Rede neural artificial

Para gerar os modelos de RNA foi utilizada a função Multilayer Perceptron disponível no WEKA, com algoritmo de aprendizado backpropagation. Na Figura 2 é apresentado um esquema da arquitetura de uma rede Multilayer Perceptron (MLP), sendo x_i os neurônios da camada de entrada; w_{ij} os pesos entre a camada de entrada e a camada oculta, podendo ter uma ou mais camadas ocultas; w_{jk} os pesos entre a camada oculta e a camada de saída; w_0 é o limiar de ativação e u a saída da rede (LANDERAS et al., 2012).

Figura 2 – Esquema simplificado da arquitetura de uma RNA



Fonte: adaptado de Quej et al. (2017, p. 64)

Os neurônios da camada de entrada recebem os dados x_i que são multiplicados pelos pesos, w_{ij}, e transmitidos para a camada oculta, onde o valor resultante é processado através da função de ativação ϕ (WAEWSAK et al. 2014).

$$u_j = \phi\left(\sum_i^m w_{ij} \,.\, x_i\right) \tag{14}$$

Para a função de ativação é possível utilizar várias funções como: sigmóide, tangente hiperbólica, função de base radial, entre outras (BEHRANG, 2010). Neste trabalho foi utilizada a função sigmoide (Eq. 15):

$$\phi(\mu) = \frac{1}{(1+exp(-\mu))} \tag{15}$$

Onde μ é o resultado da somatória $\sum_i^m w_{ij} \,.\, x_i$.

A MLP possui o algoritmo de aprendizado backpropagation, que trabalha com o método de aprendizado supervisionado. Este tipo de aprendizado fornece exemplos de valores de saída desejados, para o ajuste dos valores de saída da rede. No caso do algoritmo backpropagation a aproximação entre o valor da saída e o valor desejado é feita pelo método do mínimo global. Procura-se o mínimo global na superfície de erro, por meio do método do gradiente descendente (ALSINA et al., 2016).

A superfície de erro possui mais de um mínimo, sendo possível que a rede neural se estabilize em um mínimo local, o que não é desejável. Para evitar que isto aconteça é importante o ajuste dos valores da taxa de aprendizado (L) e do momentum (M). A taxa de aprendizado determina a taxa com que os pesos são atualizados. Quanto maior o valor de L, mais rápido a rede vai convergir para um resultado, mas é possível que a rede se estabilize em um mínimo local. Se o L for muito pequeno, o tempo para gerar os modelos pode aumentar (GARDNER; DORLING, 1998).

O M é incluído para o caso da rede tender a um mínimo local. A cada atualização dos pesos é somado um valor proporcional à atualização de pesos anterior, aumentando a taxa de atualização, porque quanto mais próximo a rede neural se encontra de um mínimo local, menor será a taxa de atualização. Portanto, a inclusão do momentum auxilia a rede neural a sair de mínimos locais (GARDNER; DORLING, 1998).

Foram testados valores para a taxa de aprendizado (L) e o momentum (M). Na primeira etapa foram testados valores de L entre 0,000001 – 1, buscando o valor de L que fornecesse estimativas de H_G com menor rRMSE. Na segunda etapa, manteve-se fixo o valor de L, variando o parâmetro M entre 0,000001 – 0,1. Estes testes foram realizados com a combinação 1 de dados, e o melhor resultado foi aplicado às demais combinações.

Além de testar valores diferentes para os parâmetros L e M, foram testados valores diferentes para os neurônios da camada oculta. Após encontrar a rede com melhor desempenho, o número de neurônios da camada oculta foi fixado, e foi adicionada outra camada oculta, consecutiva à primeira, e testados valores para a segunda camada oculta. A escolha da rede neural com a melhor arquitetura foi feita por meio do valor de rRMSE (Eq. 16) calculado na validação dos modelos.

$$rRMSE = 100 \, . \, \frac{\left(\frac{\sum_{i=1}^{m}(H_e - H_m)^2}{m}\right)^{1/2}}{\overline{H_m}} \tag{16}$$

Onde: H_m são os valores de H_G medidos, H_e são os valores de H_G estimados, $\overline{H_m}$ é o valor médio de H_m, $\overline{H_e}$ é o valor médio de H_e e m é o número de medidas.

A Tabela 2 apresenta os números de neurônios testados para cada uma das combinações de dados de entrada.

Tabela 2 – Número de neurônios por camada para as quatro combinações de variáveis de entrada testadas para a técnica RNA

Combinação	Número de neurônios	
	Uma camada	Duas camadas
C1	5, 15, 25, 35 e 45	1ª Camada: 5 2ª Camada: 5, 15, 25, 35 e 45
C2	7, 17, 27, 37, 47	1ª Camada: 37 2ª Camada: 7, 17, 27, 37 e 47
C3	7, 17, 27, 37, 47	1ª Camada: 7 2ª Camada: 7, 17, 27, 37 e 47
C4	9, 19, 29, 39 e 49	1ª Camada: 29 2ª Camada: 9, 19, 29, 39 e 49

3.5.2 Máquina de vetor de suporte

A técnica MVS foi desenvolvida por Vapnik e colaboradores, e tem como princípio a teoria do aprendizado estatístico. A MVS pode ser utilizada para classificação, regressão e previsão (CHEN; LI, 2014). Esta técnica vem sendo aplicada em diversas áreas do conhecimento, como, por exemplo, hidrologia, estudos ambientais, computação, entre outros. A MVS se apresenta como uma opção interessante, pois tem se mostrado superior a outras técnicas estatísticas e às redes neurais artificiais (MOHAMMADI et al., 2015).

O algoritmo utilizado para gerar os modelos de MVS é o SMOreg Improved. Este algoritmo foi desenvolvido primeiramente por Platt (1998), e tinha por objetivo a classificação de dados. Smola e Scholkopf (1998) apresentaram uma extensão do algoritmo, que passou a gerar regressões. Shevade (2000) realizou adaptações no

algoritmo de Smola e Scholkopf, propondo o uso de dois valores de limiar ao invés de um, tornando o algoritmo mais eficiente.

Inicialmente a MVS era capaz de realizar apenas classificações lineares, mas com o uso das funções de Kernel se tornou possível realizar mapeamentos não lineares dos dados (PIURI et al., 2013). Para gerar os modelos de MVS foram testadas duas funções de Kernel: a polinomial (Poly Kernel) e a função de base radial (RBF Kernel).

O algoritmo SMOreg Improved possibilita o ajuste de alguns parâmetros, como parâmetro C que determina o balanceamento entre a minimização do erro de treino e a complexidade do modelo (URRACA et al., 2016) e o parâmetro γ, que é o coeficiente da função RBF Kernel. O parâmetro C é um valor ajustável para as duas funções, e foram testados diferentes valores de C com base na combinação 1 de dados de entrada.

Para definir a função de Kernel utilizada e os valores dos parâmetros (C, γ) foram comparados os valores de rRMSE (%) (Eq. 16) na validação dos modelos de MVS.

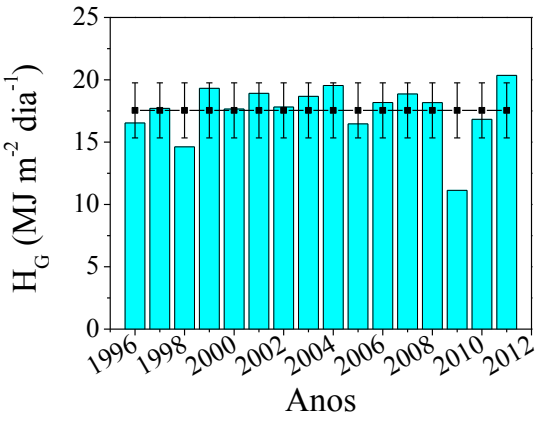
O modelo que apresentou menores valores de rRMSE foi utilizado como referência para o ajuste das combinações seguintes. A função de Kernel que apresentou melhores resultados foi a RBFKernel, com parâmetros C = 10 e γ = 1.

3.6 Validação dos modelos

Para realizar a validação foram selecionados dois bancos de dados a partir do banco de dados total, seguindo o seguinte procedimento: foram calculados os valores médios de H_G para todos os meses, de 1996 a 2011, e calculada a média inter-anos para cada mês. Os valores médios de H_G foram comparados com a média inter-anos para o respectivo mês. Os meses que apresentaram valores médios de H_G mais próximos da média inter-anos compuseram o ano típico, e os meses que apresentaram valores médios de H_G mais distantes da média inter-anos compuseram o ano atípico. O gráfico para o mês de setembro pode ser visto na Figura 3, que apresenta os valores médios de H_G para setembro de cada ano e a

média de H_G inter-anos obtida a partir de todos os dados disponíveis para este mês, com o desvio padrão.

Figura 3 – Valores médios de H_G e média de H_G inter-anos para o mês de setembro, de 1996 a 2011



A partir do gráfico é possível perceber que o valor médio de H_G para o ano de 2000 está mais próximo da média inter-anos, enquanto o valor médio de H_G para ano de 2010 está mais distante da média. Portanto para o ano típico foi selecionado o mês de setembro do ano de 2000, e para o ano atípico o mês de setembro do ano de 2010. Desta forma foram criados dois anos, de doze meses cada um, sendo que um deles era composto pelos meses com valores médios de H_G mais próximos da média de H_G inter-anos (ano típico), e o outro pelos meses com valores médios de H_G mais distantes da média de H_G inter-anos (ano atípico).

No caso dos modelos clássicos e de regressão não linear múltiplo, os anos típico e o atípico foram retirados da base de dados utilizada para a estimativa dos parâmetros dos modelos. No caso das técnicas RNA e MVS, para gerar os modelos e validar com o ano típico, foram excluídos da base de treinamento os dados correspondentes ao ano típico, e para gerar os modelos e validar com o ano atípico, foram retirados os dados correspondentes ao ano atípico.

Os indicativos estatísticos utilizados para a avaliação do desempenho das técnicas de estimativa de H_G foram: tendência do erro médio (MBE, MJ m^{-2} dia^{-1}) (Eq. 17), tendência do erro médio relativa (rMBE, %) (Eq. 18), raiz quadrada do erro

médio (RMSE, MJ m^{-2} dia^{-1}) (Eq. 19), raiz quadrada do erro médio relativa (rRMSE, %) (Eq. 16), coeficiente de correlação (r) (Eq. 20), coeficiente de determinação (R^2) (Eq. 21) e índice de concordância de Willmont (*d*) (Eq. 22).

$$MBE = \frac{\sum_{i=1}^{m}(H_e - H_m)}{m} \quad (17)$$

$$rMBE = 100 \cdot \frac{\sum_{i=1}^{m} \frac{H_e - H_m}{m}}{\overline{H_m}} \quad (18)$$

$$RMSE = \left[\frac{\sum_{i=1}^{m}(H_e - H_m)^2}{m}\right]^{1/2} \quad (19)$$

$$r = \frac{\sum_{i=1}^{m}(H_e - \bar{H}_e)\,(H_m - \bar{H}_m)}{\sum_{i=1}^{m}(H_e - \bar{H}_e)\,\sum_{i=1}^{n}(H_m - \bar{H}_m)} \quad (20)$$

$$R^2 = \frac{\left[\sum_{i=1}^{m}(H_e - \bar{H}_e)\,(H_m - \bar{H}_m)\right]^2}{\sum_{i=1}^{m}(H_e - \bar{H}_e)\,\sum_{i=1}^{n}(H_m - \bar{H}_m)} \quad (21)$$

$$d = 1 - \frac{\sum_{i=1}^{m}(H_e - H_m)^2}{\sum_{i=1}^{m}(|H_e - \bar{H}_m| + |H_m - \bar{H}_m|)^2} \quad (22)$$

Onde: H_m são os valores de H_G medidos, H_e são os valores de H_G estimados, $\overline{H_m}$ é o valor médio de H_m, $\overline{H_e}$ é o valor médio de H_e e m é o número de medidas.

Quanto menores os valores de RMSE e MBE melhor é o desempenho do modelo. Os valores de RMSE são sempre positivos, e os valores de MBE podem ser positivos ou negativos, e permitem avaliar a tendência de o modelo subestimar ou superestimar os valores de H_G, sendo que valores positivos de MBE indicam a tendência de superestimar, e valores negativos de subestimar (QUEJ et al., 2016). Quanto mais próximo de 1 estiver o r maior é a correlação entre os valores estimados e os valores medidos. O coeficiente de determinação R^2 varia entre 0 e 1, e quanto mais próximo de 1 o valor estiver melhor é o ajuste do modelo. O índice de concordância de Willmont varia de 0 a 1, e quanto mais próximo o valor estiver de 1 maior a concordância entre valores medidos e estimados.

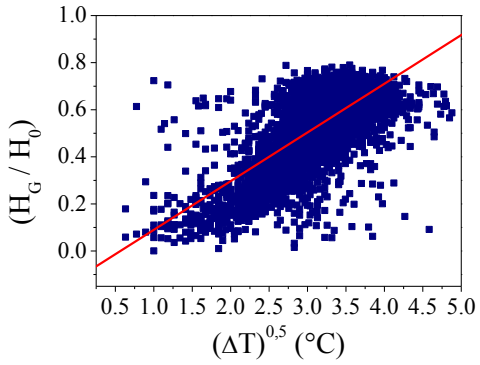
4 RESULTADOS E DISCUSSÕES

4.1 Modelos estatísticos

4.1.1 Obtenção e validação da equação HS

A Figura 4 apresenta a correlação entre a transmissividade atmosférica da irradiação solar global (H_G/H_0) e a raiz quadrada da amplitude térmica $(\Delta T_1)^{0,5}$, para um total de 4951 dias de medidas realizadas durante os anos de 1996-2011 em Botucatu – SP.

Figura 4 - Correlação entre a transmissividade atmosférica da irradiação solar global (H_G/H_0) e a raiz quadrada da amplitude térmica ($\Delta T_1^{0,5}$); a linha vermelha representa a regressão linear entre as variáveis para o período de 1996-2011



O coeficiente de calibração estimado para Botucatu foi de a_{HS} = 0,170 (Eq. 23). A Tabela 3 apresenta valores de a_{HS} ajustados para outras localidades. Mehdizadeh et al. (2016), Rahimikhoob (2010), Sanusi et al. (2013), Sharifi et al. (2016), Liu et al. (2009), Almorox et al. (2013), Almorox et al. (2011) e Almorox (2011) encontraram valores de a_{HS} menores do que o valor estimado para Botucatu. Lêdo et al. (2012), Nogueira et al. (2012), Ugwu e Ugwuanyi (2011) e Gois et al. (2016) utilizaram os valores estabelecidos anteriormente por Hargreaves de a_{HS} = 0,16 para locais distantes da costa e a_{HS} = 0,19 para locais próximos à costa ou a grandes corpos d'água. Grillone et al. (2012) e Estévez et al. (2012) encontraram intervalos de valores de a_{HS} em que o valor estimado para Botucatu está inserido.

$$\frac{H_G}{H_0} = 0{,}170.\sqrt{\Delta T_1} \qquad (23)$$

A relação entre H_G/H_0 e a umidade relativa local se reflete no valor de a_{HS}, que possui valores maiores para localidades próximas à costa, e valores menores para o interior (CITAKOGLU; 2015). Em locais próximos a grandes corpos d'água parte da energia solar é absorvida na forma de calor latente, parcela de energia utilizada para a evaporação da água, que não influencia os valores de ΔT_1, podendo fazer com que a equação tenha a tendência de subestimar H_G (LANDERAS et al.; 2012).

Tabela 3 – Valores do coeficiente de calibração a_{HS} e R² obtidos neste trabalho comparados com os valores encontrados por outros autores

Autores	Localidades	Latitude	coeficiente a_{HS}	R²	Altitude	Clima
Mehdizadeh et al. (2016)	Kerman (Irã)	30°'15	0,154		1753,8	Árido
Rahimikhoob (2010)	Ahwaz (Irã)	31° 20'	0,156	0,737	22,5	Semi-árido
Sanusi et al. (2013)	Sokoto (Nigéria)	13° 03'	0,11			
Sharifi et al. (2016)	2 (Irã)	35° 58' - 39° 46' N	0,162 e 0,165		1313 - 1364	Árido e semi-árido
//	4 (EUA)	40° - 43° N	0,136 a 0,147		378 - 1208	Semi-árido
Grillone et al. (2012)	33 Sicília (Itália)	36,96° - 38,48° N	0,146 a 0,215		40 – 1840	mediterrâneo
Estévez et al. (2012)	65 (Espanha)	36° 17' - 38° 15' N	0,160 – 0,190		4 – 1212	Semi-árido, Mediterrâneo
Liu et al. (2009)	15 (China)	34° 18' - 45° 45' N	0,141 - 0,166		3 a 2295	Clima temperado e de monção
Almorox et al. (2013)	Córdoba (Argentina)	30° 44' S	0,146			Subtropical úmido

Almorox et al. (2011)	Madri (Espanha)	40° 04' - 40° 24' N	0,146 - 0,161		477 – 604	Mediterrâneo continental temperado
Gois et al. (2016)	Seropédica (Brasil)	22° 44' S	0,16 - 0,19		26	
Lêdo et al. (2012)	Barbalha	7° 18' S	0,16		409	Tropical quente semi-árido
Nogueira et al. (2012)	Jaguaruana	4° 47' S	0,16		12	Semi-árido
Almorox (2009)	Aranjuez (Espanha)	40° 04' N	0,146			Mediterrâneo continental semi-árido
Ugwu e Ugwuanyi (2011)	Abuja (Nigéria)	8° 15' - 9° 12' N	0,16			Tropical úmido e clima seco
Modelo proposto	Botucatu	22° 53’	0,170	0,495	650 – 950	Temperado quente

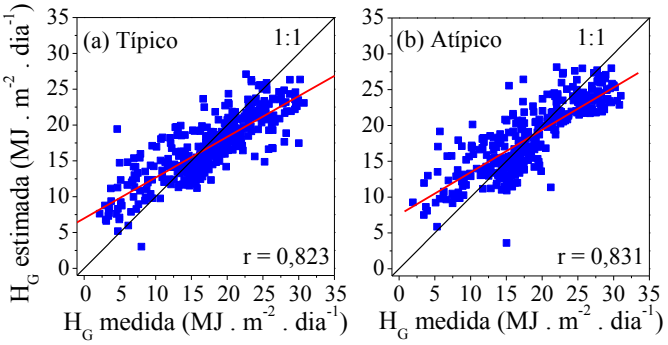
Além da umidade relativa do local, outros fatores que podem influenciar o desempenho de modelos como HS são: a latitude, a altitude, a proximidade ao mar ou a grandes corpos d’água, eventos climáticos, entre outros (LYRA et al.; 2015). Portanto é importante conhecer as condições geográficas e climáticas do local para ajustar a equação e interpretar os resultados apresentados pelo ajuste.

Rahimikhoob (2010) ajustou o modelo para Ahwaz (Irã), e o gráfico de ajuste do modelo apresentou um espalhamento mais uniforme com coeficiente de determinação de $R^2 = 0,737$, superior aos resultados deste trabalho. Rahimikhoob (2010) aplicou o modelo em uma localidade de latitude 31°20’ N, altitude de 22,5 m e clima semi-árido. Este local está situado em uma latitude maior do que a de Botucatu, de menor altitude e de clima menos úmido. Fatores climáticos e geográficos podem influenciar o desempenho de modelos baseados em valores de

ΔT_1, como dito anteriormente. Devido a variações climáticas é importante que modelos estatísticos sejam reajustados periodicamente.

A Figura 5 apresenta a correlação entre a irradiação solar global medida (H_{Gmed}) e a estimada (H_{Gest}) pelo modelo HS, para o ano típico e atípico. O coeficiente de correlação foi r_t = 0,823 e r_a = 0,831, para o ano típico e o atípico, respectivamente.

Figura 5 – Correlação entre H_G medida e H_G estimada a partir do modelo HS, para o ano típico (Fig. 5 (a)) e ano atípico (Fig. 5 (b))



A Tabela 4 apresenta os indicativos estatísticos obtidos na validação do modelo HS para os anos típico e atípico. Os valores obtidos foram: rMBEt = -3,1 % e rMBEa = 3,2 %; rRMSEt = 21,8 % e rRMSEa = 24,8 %; d_t = 0,870 e d_a = 0,880.

Tabela 4 – Indicativos estatísticos na validação do modelo HS

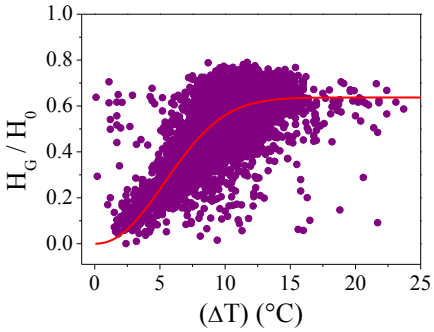
	MBE (MJ m^{-2} dia^{-1})	rMBE (%)	RMSE (MJ m^{-2} dia^{-1})	rRMSE (%)	*D*
Típico	-0,537	-3,1	3,808	21,9	0,870
Atípico	0,543	3,2	4,233	24,7	0,880

Os valores de rMBE indicam a tendência da equação de subestimar em 3,1 % os valores de H_G para o ano típico e superestimar em 3,2 % para o ano atípico, provavelmente devido a dois pontos mais distante que os demais. O valor de rRMSE menor para ano típico que para o ano atípico era esperado, em função dos valores de H_G para o ano típico serem mais próximos da média. O valor do índice de concordância de Wilmott (*d*) para o ano típico é menor do que para o ano atípico, e encontra-se dentro da normalidade em função do menor espalhamento.

4.1.2 Obtenção e validação da equação BC

A Figura 6 apresenta a correlação entre a transmissividade atmosférica da irradiação solar global (H_G/H_0) e a amplitude térmica ΔT_2. A linha em vermelho representa o modelo BC ajustado para Botucatu. O coeficiente de determinação obtido foi R^2 = 0,537. Em comparação com o modelo HS houve um pequeno aumento no valor de R², porém o ajuste do modelo BC é mais complexo, por se tratar de um ajuste não linear.

Figura 6 - Correlação entre a transmissividade atmosférica da irradiação solar global (H_G/H_0) e a raiz quadrada da amplitude térmica (ΔT_2)



A Eq. (24) apresenta o modelo BC com os coeficientes estimados para Botucatu. Os fatores que afetam o modelo BC são os mesmo que afetam o modelo HS, pois o mecanismo a partir do qual ambos os modelos estimam H_G é o mesmo. A partir dos dados de ΔT_2 é possível estimar a nebulosidade, parâmetro principal na transmissão da irradiação solar para a superfície. Assim como no modelo HS, os parâmetros da equação BC (a_{BC} , b_{BC} e c_{BC}) também precisam ser estimados para

cada local de estudo, e o modelo deve ser reajustado periodicamente, devido a variações climáticas.

$$\frac{H_G}{H_0} = 0{,}637 \cdot \left(1 - e^{\left((-0{,}014) \cdot (\Delta T_2^{2{,}177})\right)}\right) \quad (24)$$

A Tabela 5 apresenta os valores dos coeficientes de calibração (a_{BC}, b_{BC} e c_{BC}) e de RMSE no ajuste do modelo BC para outras localidades. Em relação ao parâmetro a_{BC}, os valores encontrados por Chen et al. (2004) para a China e Sharifi et al. (2016) para os EUA são similares ao valor estimado para Botucatu. Os valores encontrados por Moreno et al. (2011) para a Espanha e Sharifi et al. (2016) para o Irã são superiores ao valor encontrado neste trabalho. Em relação aos valores de b_{BC} Sharifi et al. (2016) encontraram valores similares ao valor encontrado para Botucatu. Moreno et al. (2011) e Chen et al. (2004) encontraram valores de b_{BC} maiores do que o valor encontrado para Botucatu. Moreno et al. (2011), Chen et al. (2004) e Sharifi et al. (2016) encontraram valores de c_{BC} menores do que o valor encontrado neste trabalho.

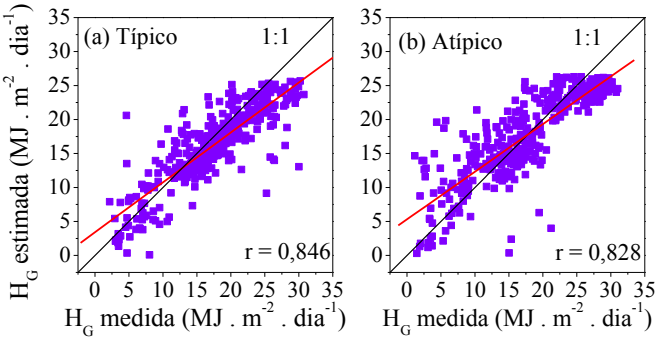
Tabela 5 – Valores dos coeficientes de calibração (a_{BC}, b_{BC} e c_{BC}) e de RMSE obtidos neste trabalho e por outros autores

Autores	Localidades	Latitude	a_{BC}	b_{BC}	c_{BC}	RMSE (MJ .m^{-2} dia^{-1})
Moreno et al. (2011)	400 (Espanha)		0,785	0,030	0,162	4,80
Chen et al. (2004)	48 (China)	18,23 - 34,72 N	0,64	0,045	1,70	
Hunt et al. (1998)	Canada	43,65 N				4,7
Sharifi et al. (2016)	2 (Irã)	35° 58' - 39° 46' N	0,879 - 1,079	0,117 - 0,158	0,636 - 0,891	3,570 - 4,150

//	4 (EUA)	40° - 43° N	0,600 - 0,673	0,012 - 0,295	1,817 - 2,101	3,075 - 3,659
Modelo proposto	Botucatu	22° 53’	0,637	0,014	2,177	3,649 – 4,138

A Figura 7 apresenta a validação do modelo BC, com os valores do coeficiente de correlação entre os valores medidos e estimados de H_G, para o ano típico (Fig. 7 (a)) e atípico (Fig. 7 (b)).

Figura 7 – Correlação entre H_G medida e H_G estimada a partir do modelo BC, para o ano típico (Fig. 7 (a)) e ano atípico (Fig. 7 (b))



Os valores do coeficiente de correlação são r_t = 0,846 e r_a = 0,828. Em comparação com os resultados para HS houve uma aumento nos valores em relação ao ano típico e uma diminuição em relação ao ano atípico.

A Tabela 6 apresenta os índices estatísticos da validação do modelo BC para os anos típico e atípico. Os valores foram: $rMBE_t$ = -7,0 % e $rMBE_a$ = 1,0 %; $rRMSE_t$ = 21,0 % e $rRMSE_a$ = 24,1 %; d_t = 0,905 e d_a = 0,901. Os valores de rMBE mostram uma tendência de o modelo subestimar os valores de H_G para o ano típico, e superestimar para o ano atípico. Os valores de rRMSE mostram um maior espalhamento nos valores de H_G atípico. O índice de concordância *d* é maior para o ano típico.

Tabela 6 – Indicativos estatísticos na validação do modelo BC

	MBE (MJ m^{-2} dia^{-1})	rMBE (%)	RMSE (MJ m^{-2} dia^{-1})	rRMSE (%)	*d*
Típico	-1,213	-7,0	3,649	21,0	0,905
Atípico	0,169	1,0	4,138	24,1	0,901

Em comparação com os resultados para o modelo HS, o valor de rMBE para o ano típico foi maior no modelo BC, e o valor para o ano atípico foi menor para este mesmo modelo, porém ambos tendem a subestimar os valores de H_G para o ano típico e superestimar para o ano atípico. No caso de HS os valores de rMBE são similares em módulo para os anos típico e atípico, e no caso de BC o valor de rMBE para o ano típico é maior do que para o ano atípico. Os valores de rRMSE para o modelo BC são menores. O índice de concordância de Wilmott para o modelo BC é maior do que para HS. A comparação entre os resultados de ambos os modelos aponta que não houve melhora significativa com a aplicação do modelo BC.

Em relação aos valores de RMSE o trabalho de Sharifi et al. (2016) apresenta valores similares aos valores do presente trabalho. Os valores encontrados por Moreno et al. (2011) e Hunt et al. (1998) são maiores do que os valores calculados para Botucatu, porém todos os resultados apresentados na Tabela 6 são da mesma ordem de grandeza que os resultados deste trabalho.

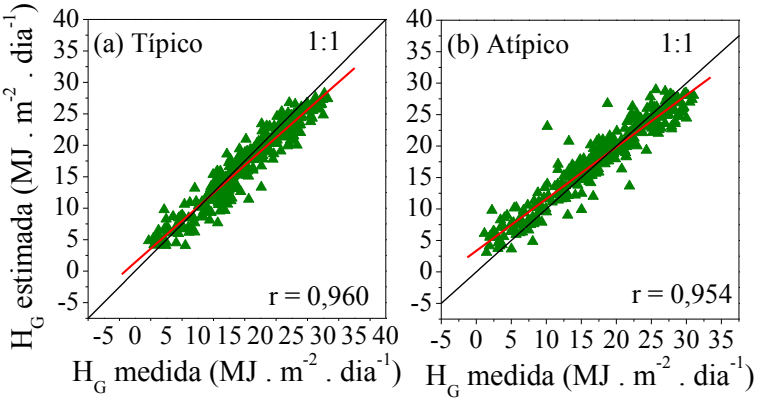
4.2 Modelo de regressão não linear múltiplo proposto

Através do método de regressão múltipla não linear buscou-se uma equação múltipla para estimar H_G, a partir dos métodos de discutidos anteriormente. A Equação (24) encontrada apresentou o melhor desempenho com base no valor de rRMSE, e foi validada utilizando a base de dados dos anos típico e atípico. Este modelo foi desenvolvido a partir da união de modelos encontrados na literatura (ANGSTRÖN, 1924; SWARTMAN; OGUNLADE, 1967; HARGREAVES, 1985).

$$H_G = 0{,}124 \,.\, H_0 \,.\, \left(\Delta {T_1}^{0{,}325}\right) + 0{,}308 \,.\, H_0 \,.\, (RI^{0{,}767}) \,.\, (UR^{0{,}075}) \quad (25)$$

A Figura 8 apresenta os gráficos da correlação entre H_G medida e H_G estimada a partir da Eq. (25), para o ano típico (Fig. 8 (a)) e atípico (Fig. 8 (b)).

Figura 8 – Correlação entre H_G medida e H_G estimada a partir da Eq. (25), para o ano típico (Fig. 8 (a)) e ano atípico (Fig. 8 (b))



A Tabela 7 apresenta os indicativos estatísticos do desempenho da equação selecionada.

Tabela 7 – Indicativos estatísticos de validação da Eq. (24)

	MBE (MJ m⁻² dia⁻¹)	rMBE (%)	RMSE (MJ m⁻² dia⁻¹)	rRMSE (%)	*d*
Típico	-0,670	-3,8	1,966	11,3	0,974
Atípico	0,339	2,0	2,324	13,6	0,971

Com base nos valores de rRMSE – $rRMSE_t$ = 11,3 % para o ano típico, e $rRMSE_a$ = 13,6 % para o ano atípico – é possível observar que a equação gera maior espalhamento no ano atípico. Com base nos valores de rMBE, para o ano típico, a equação subestima em 3,8 % o valor de H_G, e para o ano atípico a equação superestima os resultados em 2,0 %. O índice de concordância de Wilmott para os dois casos (d_t = 0,974 para o ano típico, e d_a = 0,971 para o ano atípico) são similares.

Na comparação com resultados da literatura, Swartman e Ogunlade (1967) desenvolveram quatro equações de regressão utilizando dados de razão de insolação e umidade relativa para estimar H_G para três localidades de Nigéria, e tiveram os resultados de r entre 0,805 e 0,817. A equação apresentada neste trabalho obteve resultados maiores para r, sendo possível observar a contribuição de ΔT_1 para melhorar a estimativa. Bilgili e Ozgoren (2011) desenvolveram cinco equações de regressão para estimar H_G para Adana, região mediterrânea da Turquia, utilizando as variáveis: razão de insolação, temperatura atmosférica, velocidade do vento, mês do ano e dia do mês. A primeira equação possui apenas uma variável, razão de insolação, e outras variáveis são adicionadas a cada equação desenvolvida. Os valores de r variam de 0,868 a 0,926, sendo o maior resultado para a quarta equação que não possui a variável dia do mês. Os valores de r encontrados para Botucatu são superiores.

Almorox (2011) propôs uma equação com as seguintes variáveis: radiação solar no topo da atmosfera, pressão de vapor de saturação a temperatura mínima, pressão de vapor de saturação a temperatura máxima, dados de precipitação transformados, umidade relativa mínima diária. A localidade é Aranjuez (Espanha), e os resultados encontrados foram R^2 = 0,92, RMSE = 2,37 MJ m^{-2} dia^{-1}, MBE = 0,09 MJ m^{-2} dia^{-1}. O valor de R^2 foi similar aos valores encontrados para este trabalho – 0,922 típico, e 0,910 para o ano atípico – o valor de RMSE para Aranjuez foi maior do que o encontrado para o ano típico (1,966 MJ m^{-2} dia^{-1}), e similar ao valor para o ano atípico (2,324 MJ m^{-2} dia^{-1}), e o resultado de MBE para Aranjuez foi menor, em módulo, do que os valores encontrados para Botucatu – -0,670 MJ m^{-2} dia^{-1} para o ano típico, e 0,339 MJ m^{-2} dia^{-1} para o ano atípico.

4.3 Aprendizagem de máquinas

4.3.1 Treinamento e validação das técnicas de aprendizado de máquinas

Após a realização de diversos testes variando o número de neurônios em redes neurais de uma e duas camadas ocultas, e variando os valores de L e M, encontrou-se a melhor arquitetura para as redes. A Tabela 8 apresenta número de camadas, número de neurônios por camada e os valores de L e M para cada combinação de

variáveis de entrada, e a Tabela 9 apresenta o número de épocas e os valores de rRMSE de treinamento.

Tabela 8 – Número de camadas, número de neurônios por camada e os parâmetros L e M para cada combinação da RNA

Combinação	Número de neurônios	Parâmetros
RNA1	1ª camada: 5; 2ª camada: 5	L=0,1, M=1.10^{-5}
RNA2	1ª camada: 37; 2ª camada: 27	L=0,1, M=1.10^{-5}
RNA3	1ª camada: 7; 2ª camada: 7	L=0,1, M=1.10^{-5}
RNA4	1ª camada: 29; 2ª camada: 29	L=0,1, M=1.10^{-5}

No caso da técnica MVS foram testadas duas funções de KERNEL – Poly Kernel e RBF Kernel – e diversos valores de C e γ, sendo que a arquitetura que apresentou menor valor de rRMSE foi escolhida para gerar as estimativas de H_G. A função que apresentou melhores resultados foi a RBF Kernel, com C=10 e γ=1. A Tabela 9 apresenta os valores de rRMSE para a MVS no treinamento.

Tabela 9 – Coeficientes de parametrização e resultados de rRMSE para o treinamento das técnicas de AM

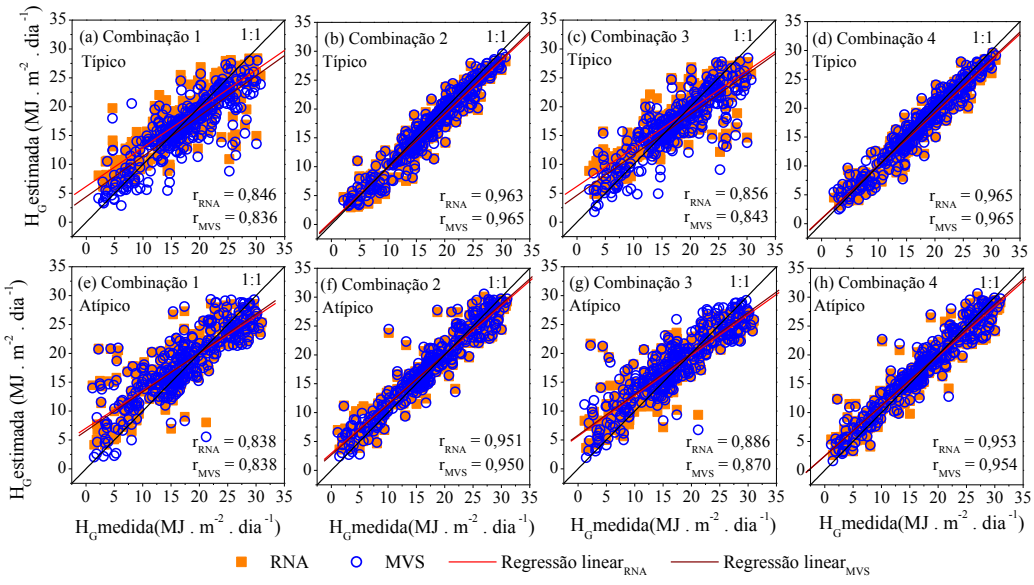
	RNA			MVS		
Combinação	Banco de dados	Épocas	rRMSE (%)	C	γ	rRMSE (%)
Combinação 1	Típico	20	21,760	10	1	20,056
	Atípico	20	19,668	10	1	19,873
Combinação 2	Típico	20	12,155	10	1	12,114
	Atípico	20	11,997	10	1	11,925
Combinação 3	Típico	20	19,755	10	1	18,918
	Atípico	20	18,610	10	1	18,970
Combinação 4	Típico	20	11,873	10	1	11,749
	Atípico	20	11,745	10	1	11,550

A partir dos dados de treinamento é possível avaliar que, para as duas técnicas, o melhor desempenho foi da combinação 4, e o pior desempenho foi da combinação 1. Comparando a combinação 1 com a 3 percebe-se uma pequena melhora devido a inclusão de UR, porém a contribuição foi menor em comparação com a combinação 2, em que RI foi incluída, reduzindo 10 % o valor de rRMSE.

Na Figura 9 encontram-se os gráficos de correlação entre H_G medida e H_G estimada para as quatro combinações de variáveis de entrada, para o ano típico ((a), (b), (c) e (d)) e atípico ((e), (f), (g) e (h)), para as técnicas RNA e MVS.

Comparando os resultados obtidos na combinação 1 com as os modelos HS e BC, com as mesmas variáveis (H_0, ΔT e H_G), as técnicas de AM apresentaram valores maiores de r, com r_t = 0,846 para o ano típico e r_a = 0,838 para o ano atípico para a RNA, e r_t = 0,836 para o ano típico e r_a = 0,838 para o ano atípico para a MVS do que os modelos clássicos (r_t = 0,823 e r_a = 0,831 para HS e r_t = 0,846 e r_a = 0,828 para BC).

Figura 9 – Correlação entre H_G medida e H_G estimada através das técnicas RNA e MVS para o ano típico (fig. 9 (a), (b), (c) e (d)) e para o ano atípico (fig. 9 (e), (f), (g) e (h))



Com a inclusão de RI houve um aumento considerável dos valores de r para a combinação 2 para as duas técnicas (r_t = 0,963 e r_a = 0,951 para RNA, r_t = 0,965 e r_a = 0,950 para MVS). Similarmente, com a inclusão de UR, combinação 3, também

houve aumento nos valores de r, mas com contribuição menor: r_t = 0,856 e r_a = 0,886 para RNA, e r_t = 0,843 e r_a = 0,870 para MVS.

Comparando os resultados da combinação 4 – r_t = 0,965 e r_a = 0,953 para a RNA e r_t = 0,965 e r_a = 0,954 para a MVS – com os resultados da RM, observa-se que para o ano típico as técnicas de AM foram superiores a RM, mas para o ano atípico os resultados foram similares. Entre a combinação 2 e a 4 a diferença de r é da ordem de centésimos, expondo a baixa correlação apresentada entre H_G e UR. A partir dos valores de r não é possível avaliar qual das técnicas, RNA ou MVS, possui o melhor desempenho.

A Tabela 10 apresenta os indicativos estatísticos de validação MBE, rMBE, RMSE, rRMSE e *d* de Wilmott para as quatro combinações e para as duas técnicas de AM.

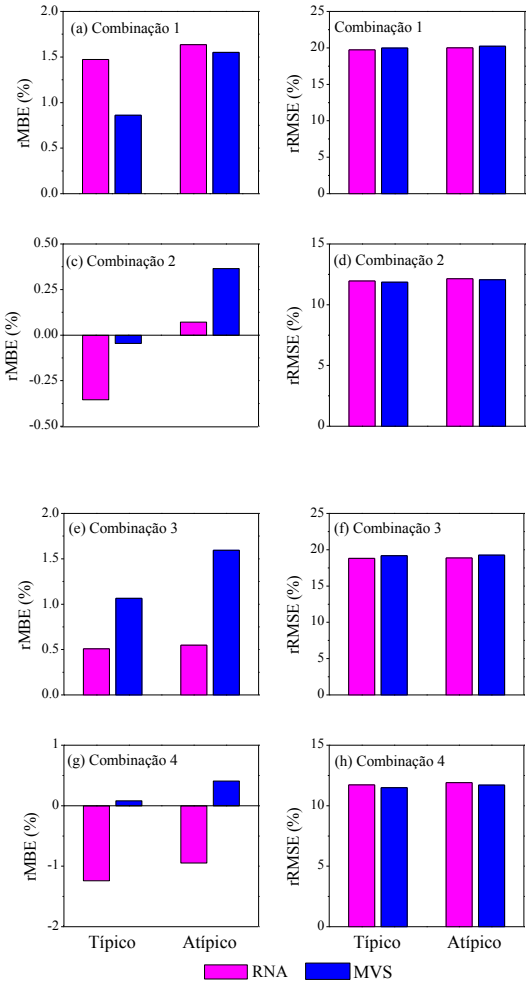
Tabela 10 – Indicativos estatísticos de validação para as técnicas RNA e MVS

Técnica	Banco de dados	MBE (MJ m^{-2} dia^{-1})	rMBE (%)	RMSE (MJ m^{-2} dia^{-1})	rRMSE (%)	*d*
RNA1	Típico	0,257	1,5	3,446	19,7	0,896
	Atípico	0,285	1,6	3,494	20,0	0,895
RNA2	Típico	-0,062	-0,3	2,088	12,0	0,967
	Atípico	0,012	0,1	2,117	12,1	0,967
RNA3	Típico	0,089	0,5	3,284	18,8	0,908
	Atípico	0,096	0,5	3,292	18,9	0,909
RNA4	Típico	-0,216	-1,2	2,047	11,7	0,969
	Atípico	-0,165	-1,0	2,078	11,9	0,968
MVS1	Típico	0,150	0,9	3,491	20,0	0,897
	Atípico	0,270	1,5	3,530	20,2	0,897
MVS2	Típico	-0,008	0,0	2,069	11,8	0,969
	Atípico	0,064	0,4	2,105	12,1	0,968
MVS3	Típico	0,186	1,1	3,348	19,2	0,907

	Atípico	0,278	1,6	3,362	19,3	0,908
MVS4	Típico	0,014	0,1	2,008	11,5	0,970
	Atípico	0,071	0,4	2,042	11,7	0,970

Na combinação 1 para a técnica RNA, os valores de rMBE foram de 1,5 % para o ano típico e 1,6 % para o ano atípico, enquanto que para a MVS foram de 0,9% para o ano típico e 1,5 % para o ano atípico. O desempenho da técnica MVS foi superior ao da RNA. Os valores de rMBE das redes foram menores que os valores dos modelos HS e BC. Os resultados de rRMSE para a RNA foram 19,7 % para o ano típico e 20,0 % para o ano atípico, enquanto que para a MVS foram 20,0 % para o ano típico e 20,2 % para o ano atípico. No geral houve uma pequena melhora nos indicativos estatísticos em relação aos modelos estatísticos, principalmente em relação ao ano atípico.

Figura 10 – Indicativos estatísticos rMBE e rRMSE das técnicas de AM, para o ano típico e atípico, para as quatro combinações de variáveis de entrada



Para a combinação 2, os valores de rMBE RNA foram de -0,3 % para o ano típico e 0,1 % para o ano atípico, e para a MVS foram de $rMBE_t$ = 0,0 % e $rMBE_a$ = 0,4 %. Os resultados apontam uma melhora no desempenho em relação a combinação 1 com a inclusão de RI. Em relação aos valores de rRMSE os valores para RNA foram de $rRMSE_t$ = 12,0 % e $rRMSE_a$ = 12,1 %, e para MVS foram de $rRMSE_t$ = 11,8 % e $rRMSE_a$ = 12,1 %. O desempenho das técnicas de AM na combinação 2 foi aproximadamente 8 % superior em relação a combinação 1 para as duas condições de validações.

Para a combinação 3 os valores de rMBE para a RNA foram de $rMBE_t$ = 0,5 % e $rMBE_a$ = 0,5 %, e para a MVS foram de $rMBE_t$ = 1,1 % e $rMBE_a$ = 1,6 %, mostrando valores maiores do que no caso da combinação 2. Os valores de rRMSE – típico 18,8 % e atípico 18,9 % para a RNA, típico 19,2 % e atípico 19,3 % para a MVS – foram maiores aos da combinação 2 e indica que a substituição de RI por UR não é vantajosa.

Para a combinação 4 os resultados para rMBE foram: -1,2 % típico e -0,9 % atípico para a RNA, típico 0,1 % e atípico 0,4 % para MVS, enquanto que os valores de rRMSE foram: típico 11,7 % e atípico 11,9 % para a RNA, e típico 11,5 %, e atípico 11,7 % para a MVS. O desempenho da RNA e MVS em estimar H_G são similares nas combinações 2 e 4.

4.3.2 Comparação do desempenho das técnicas de AM com resultados da literatura

A Tabela 11 apresenta os valores dos indicativos estatísticos, r, MBE, rMBE, R^2, RMSE e rRMSE para as técnicas RNA e MVS para diversas localidades do mundo e também os resultados para Botucatu. Para a RNA os resultados de RMSE encontrados por Lyra et al. (2015) para Alagoas, Mejdoul et al. (2013) para Marrocos e Waewasak et al. (2014) para Bangckok (Tailandia) foram superiores aos resultados para Botucatu. Os resultados encontrados por Wan et al. (2008) para a China, Landeras et al. (2012) para o País Basco, Elizondo et al. (1994) para os Estados Unidos, Bocco et al. (2011) para Salta (Argentina), Bocco et al. (2006) para Córdoba (Argentina), Linarez-Rodriguez et al. (2011) para Andaluzia (Argentina) são

similares aos resultados encontrados neste trabalho. Os resultados encontrados por Fortin et al. (2008) para Montreal (Canadá) foram inferiores aos resultados encontrados neste trabalho.

Tabela 11 – Indicativos estatísticos de estimativas realizadas a partir das técnicas RNA e MVS para diversas localidades do mundo e para Botucatu

Autores	Localidades	Modelos	MBE	rMBE	RMSE	rRMSE	r	R^2
	N'	Técnica	MJ m^{-2} dia^{-1}	%	MJ m^{-2} dia^{-1}	%		
Lyra et al. (2015)	Alagoas (Brasil)	RNA			1,97	10,1		0,75
Fortin (2008)	Montreal (Canadá)	RNA			3,83 a 5,45			
Wan et al. (2008)	China	RNA	-3,08 a 2,14	-16,0 a 18,9	1,27 a 4,00	8,2 a 24,3		
Bosch et al. (2008)	Granada (Espanha)	RNA		-1,2 a 2,1		5,0 a 7,5		0,98
Mejdoul et al. (2013)	8 (Marrocos)	RNA			1,23 a 1,5		0,98	
Landeras et al. (2012)	País Basco (Espanha)	RNA			2,93			
Waewasak et al. (2014)	Bangckok (Tailândia)	RNA	-0,02 – 0,003		0,003 a 0,392			
Elizondo et al. (1994)	3 (Estados Unidos)	RNA			2,92 a 3,64			
Bocco et al. (2006)	Córdoba	RNA			3,14 a 3,88		0,86 a 0,92	0,96 a 0,99
Linarez-Rodríguez et al. (2011)	Andaluzia (Espanha)	RNA			2,88	16,4	0,94	
Urraca et al. (2015)	4 (Espanha)	RNA			2,11 a 2,9			0,90 a 0,93
Modelo proposto	Botucatu	RNA	-0,216	-1,240	2,047	11,7 29	0,969	0,939
Modelo proposto	Botucatu	RNA	-0,165	-0,946	2,078	11,9 10	0,968	0,937

Ramedani et al. (2014)	Teerã (Irã)	MVS			3,30 a 3,40			0,887 a 0,889
Chen et al. (2015)	32 (China)	MVS			0,840 a 2,660			
Chen et al. (2013)	3 (China)	MVS			1,79 a 2,38			
Quej et al. (2017)	Yucatán (México)	MVS			2,578 a 2,877			
Mohammadi et al. (2016)	Bandar Abbas (Irã)	MVS			2,0832 a 2,2794			0,9173 a 0,9380
Modelo proposto	Botucatu	MVS	0,014	0,083	2,008	11,504	0,970	0,941
Modelo proposto	Botucatu	MVS	0,071	0,409	2,042	11,708	0,970	0,941

Em relação aos valores de rMBE encontrados por Bosch et al. (2008) para Granada (Espanha) foram similares aos resultados para Botucatu, e os resultados encontrados por Wan et al. (2008) para a China foram inferiores. Bocco et al. (2006) e Linarez-Rodriguez et al. (2011) encontraram valores de r similares aos encontrados neste trabalho, e Mejdoul et al. (2013) encontrou valores superiores. Em relação aos valores de R^2 os resultados encontrados por Lyra et al. (2015) foram inferiores aos resultados encontrado neste trabalho. Os resultados encontrados por Bosch et al. (2008) foram superiores, e os resultados encontrados por Bocco et al. (2010) e Bocco et al. (2006) foram similares aos resultados encontrados neste trabalho.

Para a técnica MVS os valores de RMSE encontrados por Chen et al. (2015) para a China apresentam intervalo de valores no qual os valores encontrados para Botucatu estão inseridos. Os valores encontrados por Ramedani et al. (2014) para Teerã (Irã), Chen et al. (2013) para a China, Quej et al. (2015) para o México, Quej et al. (2017) para Yucatán (México), Mohammadi et al. (2015) para Bandar Abbas (Irã) foram similares aos resultados para Botucatu. Em relação aos valores de R^2, Ramedani et al. (2014) e Mohammadi et al. (2015) encontraram resultados similares aos resultados para Botucatu.

4.3.3 Comparação do desempenho das técnicas de AM com os modelos estatísticos

A Tabela 12 apresenta os resultados de r e rRMSE para os modelos estatísticos HS e BC e para as técnicas RNA e MVS no caso da combinação 1, as quais possuem as mesmas variáveis de entrada. Em relação aos valores de r, as técnicas de AM melhoraram os resultados, sendo que para o ano típico a técnicas RNA apresentou melhor resultado, e para o ano atípico as duas técnicas apresentaram resultados iguais.

Tabela 12 – Valores de rRMSE e r para os modelos estatísticos e para as técnicas de AM para a combinação 1

	Banco de dados	rRMSE (%)	R
HS	Típico	21,9	0,823
	Atípico	24,7	0,831
BC	Típico	21,0	0,846
	Atípico	24,1	0,828
RNA1	Típico	19,7	0,846
	Atípico	20,0	0,838
MVS1	Típico	20,0	0,836
	Atípico	20,2	0,838

O desempenho em estimar H_G das técnicas de AM e os modelos estatísticos HS e BC são similares estatisticamente, onde os valores de rRMSE e r são aproximadamente iguais quando as variáveis de entrada são as mesmas.

4.3.4 Comparação do desempenho das técnicas de AM com a equação de RM.

A Tabela 13 apresenta os valores de r e rRMSE para a equação RM e as técnicas de AM combinação 4. Em relação aos valores de r e rRMSE não há diferença significativa entre os resultados de RNA, MVS da RM, com exceção do valor de rRMSE da equação RM na condição de validação atípica a qual foi maior em aproximadamente 2,5%.

Tabela 13 – Valores de rRMSE e r para o modelo de regressões múltiplas e para as técnicas de AM para a combinação 4

	Banco de dados	rRMSE (%)	R
RM	Típico	11,3	0,960
	Atípico	13,6	0,954
RNA4	Típico	11,7	0,965
	Atípico	11,9	0,953
MVS4	Típico	11,5	0,965
	Atípico	11,7	0,954

O desempenho em estimar H_G das técnicas de AM na combinação 2 e 4, e os modelo estatístico RM, são similares estatisticamente, mesmo quando as variáveis de entrada são as mesmas, com exceção da RM na validação em condições atípicas.

5 CONCLUSÕES

Dos resultados apresentados e discutidos sobre os modelos de estimativa de H_G diária: estatístico de HS**,** BC e RM, e por técnicas de aprendizado de máquinas RNA e MVS, propostos para a cidade de Botucatu podemos concluir:

O coeficiente da equação de Hargreaves-Samani a_{HS} = 0,170 obtido em Botucatu SP/Brazil, depende do clima, latitude e altitude, e pode der utilizada na estimativa de H_G com restrições devido ao baixo coeficiente de determinação (R^2 = 0,495). Na validação do modelo HS, comparação, entre H_G estimada e H_G medida em Botucatu, os valores obtidos foram: coeficiente de correlação r_t = 0,823 e r_a = 0,831; os indicativos estatísticos $rMBE_t$ = -3,1 % e $rMBE_a$ = 3,2% ; $rRMSE_t$ = 21,9 % e $rRMSE_a$ = 24,7%, para o ano típico e o atípico respectivamente indicam que o modelo HS pode ser utilizado na estimativa de H_G com regular precisão e exatidão.

O coeficiente não linear do modelo de BC obtido a = 0,637 na cidade de Botucatu, SP/Brazil, depende do clima, latitude e altitude e pode der utilizada na estimativa de H_G com restrições devido ao baixo coeficiente de determinação (R^2 = 0,537).

Na validação da equação de BC, os valores obtidos: coeficiente de correlação r_t = 0,846 e r_a = 0,828 ; e os indicativos estatísticos $rMBE_t$ = -7,0 % e $rMBE_a$ = 1,0 %; $rRMSE_t$ = 21,0 % e $rRMSE_a$ = 24,1 % para o ano típico e o atípico, respectivamente, indicando que o modelo BC pode ser utilizado na estimativa de H_G com regular precisão e exatidão.

O modelo de regressão múltipla desenvolvido neste trabalho apresentou indicativos de validação: r_t = 0,960 e r_a = 0,956 ; $rMBE_t$ = - 3,8 % , $rMBE_a$ =2,0 %; rRMSEt = 11,3 %, rRMSEa = 13,6 % para o ano típico e o atípico, respectivamente.

As técnicas MVS e RNA, com as mesmas variáveis de entrada (H_0 e $\Delta T^{0,5}$) que o modelo HS, mostram por meio dos valores dos indicativos estatísticos: MVS ($rMBE_t$ = 0,9% e $rMBE_a$ = 1,6 %; $rRMSE_t$ = 20,0 % e $rRMSE_a$ = 20,2 %) e RNA ($rMBE_t$ = 1,4% e $rMBE_a$ = 1,6 %; $rRMSE_t$ = 19,7 % e $rRMSE_a$ = 20,0 %) para o ano típico e o atípico, respectivamente, e podem estimar H_G com desempenho similar aos modelos estatísticos de HS e BC.

Os valores crescentes dos valores de r com entradas de diferentes variáveis meteorológicas mostram que as técnicas RNA e MVS elevam o desempenho dos modelos AM nas estimativas de H_G.

As técnicas MVS e RNA na combinação 4, com as entradas das variáveis meteorológicas: irradiação solar global no topo da atmosfera (H_0, MJ m-2 dia-1); e temperatura máxima- ($T_{máx}$, °C); temperatura mínima (T_{min}, (°C); razão de insolação (RI) e umidade relativa do ar (UR, %), com valores dos coeficientes de correlações r_t= 0,960 e r_a=0,95, e valores dos indicativos estatísticos: MVS ($rMBE_t$ = 0,1 % e $rMBE_a$ = 0,4 %; $rRMSE_t$ = 11,5 % e $rRMSE_a$ = 11,7 %); e RNA($rMBE_t$ = -1,2 % e $rMBE_a$ = -1,0 %; $rRMSE_t$ = 11,7 % e $rRMSE_a$ = 11,9 %), para o ano típico e o atípico, respectivamente, podem estimar H_G com melhor desempenho que as modelos de HS e BC, e desempenho similar ao modelo RM.

A comparação entre os desempenhos dos modelos HS, BC, RM e as redes MVS e RNA na combinação 4, sugere o uso do modelo RM, quando os dados de RI e UR estejam disponíveis, por ser um método mais prático que o uso das técnicas de aprendizado de máquinas.

REFERÊNCIAS

ALLEN, R. G. Self-Calibrating Method for Estimating Solar Radiation from Air Temperature. **Journal of Hidrologic Engineering**, v. 2, n. 2, p. 56 – 67, 1997.

ALMEIDA, W. G. **Avaliação de um modelo físico estimadorde irradiância solar baseado em satélites geoestacionários.** 1998. Dissertação (Mestrado em Engenharia Mecânica) – Universidade Federal de Santa Catarina, Florianópolis, 1998.

ALMOROX, J. Estimating global solar radiation from common meteorological data in Aranjuez, Spain. **Turkish Jornal of Physics**, v. 35, p. 53-64, 2011.

ALMOROX, J.; BENITO, C. H. Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain). **Applied Energy**, v. 88, p. 1703-1709, 2011.

ALMOROX, J.; BOCCO, M.; WILLINGTON, E. Estimation of daily global solar radiation from measured temperatures at Cañada de Luque, Córdoba, Argentina. **Renewable Energy**, v. 60, p. 382-387, 2013.

ALSINA, E. et al. Artificial neural network optimisation for monthly average daily global solar radiation prediction. **Energy Conversion and Management**, v. 120, p. 320-329, 2016.

ALVARENGA, L. A. Precipitação no sudeste brasileiro e sua relação com a Zona de Convergência do Atlântico Sul. **Revista Agrogeoambiental**, 2012. DOI: http://dx.doi.org/10.18406/2316-1817v4n22012452

ANNANDALE, J. G. et al. Software for missing data error analysis of Penman-Monteith reference evapotranspiration. **Irrigation Science**, 2002. DOI 10.1007/s002710100047.

ANGSTRÖM, A. Solar and terrestrial radiation. **Quartely Journal of the Royal Meteorological Society**, v. 50, p. 121-125, 1924.

ANTONANZAS-TORRES, F. et al. Generation of daily global solar irradiation with support vector machines for regression. **Energy Conversion and Management**, v. 96, p. 277-286, 2015.

AYODELE, T. R.; OGUNJUYIGBE, A. S. O.; MONYEI, C. G. On the global solar radiation predictions methods. **Journal of Renewable and Sustainable Energy**, v. 8, 2016. DOI: 10.1063/1.4944968

BECHINI, L.; DUCCO, G.; DONATTELI, M.; STEIN, A. Modelling, interpolation and stochastic simulation is space and time of global solar radiation. **Agriculture, Ecosystems and Environment**, 2000. https://doi.org/10.1016/S0167-8809(00)00170-5.

BEHRANG et al. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. **Solar Energy**, v. 84, p.1468 - 1480, 2010.

BELAID, S.; MELLIT, A. Prediction of daily and mean monthly global solar radiation using support vector machine in na arid climate**. Energy Conversion and Management**, v. 118, p. 105 – 118, abril, 2016.

BERRY, W. D.; FELDMAN, S. **Multiple regression in practice:** Series: Quantitative Applications in the Social Sciences. California: Sage Publications, 1985. p. 1-5.

BESHARAT F.; DEHGHAN, A. A.; FAGHIH A. R. Empirical models for estimating global solar radiation: A review and case study. **Renewable and Sustainable Energy Reviews**, v. 21, p. 798-821, 2013.

BILGILI, M.; OZFOREN, M. Daily total global solar radiation modeling from several meteorological data**. Meteorology and Atmospheric Physics**, 2011. DOI 10.1007/s00703-011-0137-9.

BOCCO, M.; OVANDO, G.; SAYAGO, S. Devolopment and evaluation of neural network models to estimate daily solar radiation at Córdoba, Argentina. **Pesquisa Agropecuária Brasileira**, v. 41, p. 179-184, 2006.

BOCCO, M.; WILLINGTON, E.; ARIA, M. Comparison of regression and neural networks models to estimate solar radiation. **Chilean Journal of Agricultural Reseach**, v. 70, n. 3, p. 428-435, 2010.

BOJANOWSKI, J. S.; DONATELLLI, M.; SKIDMORE, A. K.; VRIELING, A. Na auto-calibration procedure for empirical solar radiation models. **Environmental Modelling and Software**, 2013. https://doi.org/10.1016/j.envsoft.2013.08.002.

BONACCORSO, G. **Machine Learning Algorithms**: A reference guide to popular algorithms for data science anda machine learning. Birmingham: Packt Publishing, 2017, cap. 1, p. 8.

BOSCH, J. L.; LÓPEZ, G.; BATLLES, F. J. Daily solar irradiation estimation over a mountainous area using artificial neural networks. **Renewable Energy**, v. 33, p. 1622-1628, 2008.

BRISTOW, K. L.; CAMPBELL, G. S. On the relationship between incoming solar radiation and daily maximum and minimum temperature. **Agricultural and Forest Meteorology**, v. 31, n. 2, p. 159 – 166, maio, 1984.

CHAVES, M. A.; ESCOBEDO, J. F. A software to process daily solar radiation data. **Renewable Energy**, v. 19, p. 339 – 344, 2000.

CHAVES, M. A.; ESCOBEDO, J. F.; TEODORO, S. M. “Horas de brilho solar e índice de insolação: uma comparação entre os valores medidos com heliógrafo e pireliômetro”. In: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 2, 1999, Florianópolis, SC: **Epagri**, 1999. CD-ROM.

CHEN, R. et al. Validation of five global radiation models with measured daily data in China. **Energy Conversion and Management**, v. 45, p. 1759-1769, 2004.

CHEN, J. L; LI, G. S.; WU, S. J. Assessing the potentical of support vector machine for estimatting daily solar radiation using Sunshine duraction. **Energy Conversion and Management**, v. 75, p. 311-318, 2013.

CHEN, J-L.; LI, G-S. Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. **Theoretical and Applied Climatology**, v. 115, n. 3 – 4, p. 627 – 638, fevereiro, 2014.

CHEN et al. Assessing the transferability os support vector machine model for estimation of global solar radiation from air temperature. **Energy Conversion and Management**, v. 89, p. 318-329, 2015.

CITAKOGLU, H. Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. **Computers and Electronics in Agriculture**, v. 118, p. 28-37, 2015.

CROTTI, P.; RAMPINELLI, G. A. Estimativa das componentes direta e difusa da radiação solar em superfície horizontal para Araranguá/SC a partir de redes meteorológicas. In: COGRESSO BRASILEIRO DE ENERGIA SOLAR, 7., 2018, Gramado. **Anais** [...]. Gramado, 2018. Disponível em: https://anaiscbens.emnuvens.com.br/cbens/issue/view/1

DONATELLI, M.; CAMPBELL, G. S. A simple model to estimate global solar radiation. **Procedings of the 5th European Sociaty of Agronomy Congress**. Nitra, Slovak Republic; p. 133-134, 1998.

ECHER, E.; SOUZA, M. P.; SCHUCH, N. J. A Lei de Beer Aplicada na Atmosfera Terrestre. **Revista Brasileira de Ensino de Física**, v. 23, n. 3, p. 276-283, 2001.

ELIZONDO, D.; HOOGENBOOM, G.; MCCLENDON, R. W. Development of a neural network model to predict daily solar radiation. **Agricultural and Forest Meteorology**, v. 71, p. 115-132, 1994.

ERTEKIN, C.; YALDIZ, O. Estimation of monthly average daily global radiation on horizontal surfasse for Antalya (Turkey). **Renewable Energy**, v. 17, p. 95-102, 1999.

ESTÉVEZ, J.; PADILLA, F. L. M.; GAVILÁN, P. Evaluation and Regional Calibration of Solar Radiation Prediction Models in Southern Spain. **Journa of Irrigation and Drainage Enginnering**, v. 138, n. 10, p. 868-879, 2012.

FALAYI, E. O.; ADEPITAN, J. O.; RABIU, A. B. Empirical models for the correlation of global solar radiation with meteorological data for Isayin, Nigeria. **International Journal of Physical Sciences**, v. 3, n. 9, p. 210-216, 2008.

FENG, Y. et al. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. **Atmosphere Environment**, v. 45, n. 11, p. 1979 – 1985, abril, 2011.

FERRARI, S. et al. "Computational intelligence models for solar radiation prediction," ***2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)***, Minneapolis, MN, 2013, p. 757-762, doi: 10.1109/I2MTC.2013.6555517.

FIORIN, D. V. et al. Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. **Revista Brasileira de Ensino de Física**, v. 33, n. 1, p. 1- 20, 2011.

FORTIN et al. Comparision of empirical daily surface incoming solar radiation models. **Agricultural and Forest Meteorology**, v. 148, 1332-1340, 2008.

GARDNER, M. W.; DORLING, S. R. Artificial Neural Networks (The Multilayer Perceptron) – A Review of Applications in the Atmospheric Sciences. **Atmospheric Environment**, v. 32, n. 14/15, p. 2627 – 2636, junho, 1998.

GARSON, G. D. **Neural Networks**: An Introductory Guide for Social Scientists. Londres: Sage Publications, 1998. cap. 1, p. 2-5.

GOIS, G. et al. Models to estimate incidente solar radiation on Seropédoca, Rio de Janeiro. **Bioscience Journal**, v. 32, n. 2, p. 505-513, 2016.

GOODIN, D. G. et al. Estimating solar irradiance for crop modeling using air temperature data. **Agronomy Journal**, v. 91, n. 5,p. 845-851, 1999.
DOI: 10.2134/agronj1999.915845x

GRILLONE, G.; AGNESE, C. D'ASARO, F. Estimation of Daily Solar Radiation from Measured Air Temperature Extremes in the Mid-Mediterranean Area. **Journal of Irrigation and Drainage Engineering**, v. 138, n. 10, p. 939-947, 2012.

HARGREAVES, G. H.; SAMANI, Z. A. Estimating Potential Evapotranspiration. **Journal of the Irrigation and Drainage Division**, v. 108, n. 3, p. 225 – 230, 1982.

HARGREAVES et al. Irrigation Water Requirements For Senegal River Basin. **Journal of Irrigation and Drainage Engineering**, v. 111, n. 3, p. 265-275, 1985.

HUNT, L. A. et al. Estimation of solar radiation for use crop modeling. **Agricultural Forest Meteorology**, v. 91, p. 293-300, 1998.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Atlas Brasileiro de Energia Solar**. São José dos Campos, 2017, 80 p.

KOTSIANTIS, S. B. Supervised Machine Learning: a Review of Classification Techniques. In: MAGLOGIANNIS, I. et al. **Emerging Artifitial Intelligence: Aplications in Computer Enginneering**. Amsterdã: IOS Press, 2007. p. 3.

KOUSKY, V. E.; GAN, M. A. Upper tropospheric cyclonic vortices in the tropical South Atlantic. **Tellus**, v. 33, n. 6, p. 538-551, 1981.

KULIK, S. D. Neural network model or artificial intelligence for handwriting recognition. Journal of Theoretical and Applied Information Technology, v. 73, n. 2, p. 202-211, 2015.

KUMAR, R. Comparison of regression and artificial neural network models for estimation of global solar radiations. **Renewable and Sustainable Energy Reviews**, v. 52, p. 1294-1299, 2015.

LANDERAS, G. et al. 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 Conversion and Management**, v. 62, p. 1-13, 2012.

LÊDO, E. R. F. et al. Avaliação de modelo de estimativa da radiação solar global (Rs) com base na amplitude térmica. **Conexões Ciência e Tecnologia**, v. 6, n. 1, p. 15-26, 2012.

LI et al. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. **Energy Conversion and Management**, v. 70, p. 139-148, 2013.

LIAKOS, K. G. et al. Machine Learning in Agriculture: a Review. **Sensors in Agriculture**, v. 18, n. 8, 2018. Disponível em: https://doi.org/10.3390/s18082674.

LINARES-RODRÍGUEZ, A. et al. Generation of synthetical daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. **Energy**, v. 36, p. 5356-5365, 2011.

LIU, X. et al. Evaluation of temperature-based global solar radiation models in China. **Agricultural and Forest Meteorology**, n. 149, p. 1433-1446, 2009.

LYRA et al. Estimation of monthly global solar irradiation using the Hargreaves-Samani model and na artificial neural network for the state of Alagoas in northeastern Brazil. **Theoretical Applied Climatology**, DOI 10.1007/s00704-015-1541-8, 2015.

MACAGNAN, M. H. **Introdução à radiação solar**. São Leopoldo, Universidade do Vale do Rio dos Sinos, 2010, 100p.

MAGHRABI, A. H. Parametrization of a simple model to estimate monthly global solar radiation based on meteorological variables, and evaluation of existing solar radiation for Tabouk, Saudi Arabia. **Energy Conversion and Management**, v. 50, p. 2754-2760, 2009.

MATTOS, T. B. **Modelos Não Lineares e suas Aplicações**. 2013. 59 f. Trabalho de conclusão de curso (Bacharel em Estatística) - Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

MEHDIZADEH, S.; BEHMANESH, J.; KHALILI, K. Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 146, p. 215 – 227, junho, 2016.

MEJDOUL, R.; TAQI, M.; BELOUAGGADIA, N. Artificial neural network based prediction model of daily global solar radiation in Morocco. **Journal of Renewable and Sustainable Energy,** 2013. DOI: 10.1063/1.4854817.

MEZA, F.;VARAS, E. Estimation of mean monthly solar global radiation as a function of temperature. **Agricultural and Forest Meteorology**, v.100, p. 231-241, 2000.

MICROCAL SOFTWARE, INC. Data analysis and technical graphics software. **Origin Professional Version 6.0**. Northampton, MA: Copyright©, 1999, 772p.

MOHAMMADI, K. et al. Extreme learning machine based prediction of daily dew point temperature. **Computers and Electronics in Agriculture**, v. 117, p. 214-225, 2015.

MOHAMMADI, K. et al. Support vector regression based prediction of global solar radiation on a horizontal surface. **Energy Conversion and Management**, v. 91, p. 433-441, 2015.

MOHAMMADI, K. et al. Temperature-based estimation of global solar radiation using soft computing methodologies. **Theoretical and Applied Climatology**, v. 125, p. 101-112, 2016.

MOHRI, M.; ROSTAMIZADEH, A; TALWALKAR, A. **Foundations of Machine Learning**, 2. ed. Cambridge: MIT Press, 2018. cap. 1, p. 1.

MOOJEN, T. M. B.; CAVALCANTE, R. B. L.; MENDES, C. A. B. Avaliação da radiação solar com base em dados de nebulosidade. **Geografia**, v. 21, n. 3, p. 41-55, 2012.

MORENO, A.; GILABERT, M. A.; MARTÍNEZ, B. Mapping daily global solar irradiation over Spain: A comparative study of selected approaches. **Solar Energy**, v. 85, p. 2072-2084, 2011.

NOGUEIRA, D. H.; SILVA, M. G.; LEDO, E. R. F.; ARAUJO, E. M. Estimativa da radiação solar global (RS) através da amplitude térmica diária. In: INOVAGRI INTERNATIONAL MEETING; WORKSHOP INTERNACIONAL DE INOVAÇÕES TECNOLÓGICAS NA IRRIGAÇÃO, 4., 2012, Fortaleza. Anais... Fortaleza: **INOVAGRI Meeting**, 2012.

PALÁCIOS, R. S. **Interação entre a radiação solar direta e os aerossóis atmosféricos na Amazônia**, 2017. Tese (Doutorado Física Ambiental) – Universidade Federal do Mato Grosso, Cuiabá, 2017

PIRI, J.; KISI, O. Modelling solar radiation reached to the Earth using ANFIS, NN-ARX, and empirical models (Case studies: Zahedan and Bojnurd stations). **Journal of Atmospheric and Solar-Terrestrial Psysics**, v. 123, p. 39 – 47, dezembro, 2014.

PIRI et al. Prediction of the solar radiation on the Earth using support vector regression technique. **Infrared Physics &Tecnology**, v. 68, p. 179-185, 2015.

PLATT, J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. **Redmond: Microsoft Research**, 1998. 21 p. TechReport.

PODESTÁ, G. P.; Núñez, L., VILLANUEVA, C. A., SKANSI, M. A. Estimating daily solar radiation in the Argentine Pampas. **Agricultural and Forest Meteorology**, 2004. https://doi.org/10.1016/j.agrformet.2003.11.002.

QUEJ, V. H.; ALMOROX, J.; IBRAKHIMOV, M.; SAITO, L. Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico. **Energy Conversion and Management**, v. 110, p. 448-456, 2016.

QUEJ, V. et al. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. **Journal of Atmospheric and Solar–Terrestrial Physics,** v. 155, p. 62 – 70, 2017.

RADHIKA, Y.; SHASHI, M. Atmospheric Temperature Prediction using Support Vector Machines. **International Journal of Computer Theory and Engineering**, v. 1, n. 1, p. 55 – 58, abril, 2009.

RAJASEKARAN, S.; GAYATHRI, S.; LEE, T-L. Support vector regression methodology for storm surge predictions. **Ocean Engineering**, v. 35, n. 16, p. 1578 – 1587, novembro, 2008.

RAHIMIKHOOK, A. Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. **Renewable Energy**, v. 35, p. 2131-3135, 2010.

RAMEDANI, Z.; OMID, M.; KEYHANI, A. A method based on neural networks for generating solar radiation map**. International Journal of Energy and Environment**, v. 3, n. 5, p. 775-786, 2012.

RAMEDANI, Z. et al. (2014). A comparative study between fuzzy linear regression and support vector regressions for global solar radiation prediction in Iran. **Solar Enery**, v. 109, p. 135-143, 2014.

RAMOS, C. M. C. et al. Modelagem da variação horária da temperatura do ar em Petrolina, PE, e Botucatu, SP. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 9, p. 959-965, 2011.

R CORE TEAM. **R: A Language and Environment for Statistical Computing** 3.6.0. Austria: R Foundation for Statistical Computing, 2018. Disponível em: https://www.r-project.org/.

RONDÓN, A.C.B.; DUCATI, J. R.; HAAG, R. ANÁLISE ANUAL DAS COMPONENTES DIRETA, DIFUSA E GLOBAL DA RADIAÇÃO SOLAR EM PORTO ALEGRE, RS. In: Congresso Brasileiro de Energia Solar, 8, 2020, Fortaleza. **Anais** [...]. Fortaleza, 2017. Disponível em: https://anaiscbens.emnuvens.com.br/cbens

SANUSI, Y. K.; ABISYE, S. G.; ABIODUN, A. O. Application of Artificial Neural Networks to Predict Daily Solar Radiation in Sokoto. **International Journal of Current Engineering and Technology**, v. 3, n. 2, p. 647-652, 2013. http://inpressco.com/category/ijcet

SANTOS, C. M**. Modelagem da irradiação direta na incidência normal em Botucatu: aprendizado de máquinas, estatístico e LINKE**. 2016. 131f. Tese (Doutorado em Agronomia/Irrigação e Drenagem) - Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Botucatu, 2016.

SHARIF, S. S.; REZAVERDINEJAD, V.; NOURANI, V. Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: A comparative study of selected temperaturebased approaches. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 149, p. 131-145, 2016.

SHEVADE et al. Improvement to the SMO Algorithm for SVM Regression. **IEEE Transactions on Neural Networks**, v. 11, n. 5, 2000.

SILVA, C. R. et al. Radiação solar estimada com base na temperatura do ar para três regiões de Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 16, n. 3, p. 281-288, 2012.

SILVA, M. B. P. **Estimativa da irradiação solar global pelo método Angström-Prescott e técnicas de aprendizado de máquinas**. 2016. 86 f. Dissertação (Mestrado em Agronomia/Irrigação e Drenagem) - Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Botucatu, 2016.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v.14, p.199-222, 2004.

SOARES, P. L. B.; SILVA, J. P. Aplicação de redes neurais artificiais em conjuntos com o método vetorial da propagação de Feixes na análise de um acoplador direcional baseado em fibra ótica. **Revista Brasileira de Computação Aplicada**, v. 3, n. 2, p. 58-72, 2011.

SONDAK, N. E.; SONDAK, V. K. Neural network and artificial intelligence. In: PROCEEDINGS OF THE 20TH SIGCSE TECHNICAL SYMPOSIUM ON COMPUTER SCIENCE EDUCATION, 1989, p. 241-245.

SWARTMAN, R. K.; OGUNLADE, O. Solar Radiation Estimates from Common Parameters. **Solar Energy**, v. 11, n. 3 – 4, p. 170 – 172, 1967.

TEIXEIRA Jr., L. A.; PESSANHA, J, F. M.; SOUZA, R. C. Análise wavelet e redes neurais artificiais na previsão da velocidade de vento. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 43, 2011, Ubatuba. **Anais:** XLIII Simpósio Brasileiro de Pesquisa Operacional, Ubatuba – Sociedade Brasileira de Pesquisa Operacional, 2011, p. 684 – 695.

TERAMOTO, E. T.; ESCOBEDO, J. F. Análise da frequência anual das condições de céu em Botucatu, São Paulo. **Revista Brasileira de Engenharia Agrícola e Ambiental,** Campina Grande, v. 16, n. 9, p. 985 – 992, 2012.

TORRES, W. L. V.; OLIVEIRA, J. B.; ARRAES, F. D. D.; CORDEIRO, A. Y. P. A. V. G.; ALVES, A.S. Avaliação de Modelos de Estimativa da Radiação Solar na Escala

Mensal em Fortaleza – CE. In: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 19, 2015, Lavras. **Anais:** agrometeorologia no século 21 - o desafio do uso sustentável dos biomas brasileiros, Lavras: Universidade Federal de Lavras – Sociedade Brasileira de Agrometeorologia, 2015, p. 73 – 81.

TYMVIOS et al. Comparative study of Angstrom and artificial neural networks' methodologies in estimating global solar radiation. **Solar Energy**, v. 78, p. 752 - 762, 2005.

UGWU, A. I.; UGWUANYI, J. U. Performance assessment of Hargreaves model in estimating solar adiation in Abuja using minimum climatological data. **International Journal of the Physical Sciences**, DOI: 10.5897/IJPS11.1403, 2011.

URRACA, R. et al. Estimation of solar global irradiation in remote areas. Journal of Renewable and Sustainable **Energy**, DOI: 10.1063/1.4919084, 2015.

VIGANÓ, H. H. G. et al. Redes Neurais Artificiais na Previsão de Queimadas e Incêndios no Pantanal. **Revista Brasileira de Geografia Física,** Recife, v. 10, n. 5, 2017.

WAEWSAK, J. et al. Estimation of Monthly Mean Daily Global Solar Radiationover Bangkok, Thailand using Artificial Neural Networks. **Energy Procedia**, v. 57, p. 1160-1168, 2014.

WAN, K. K. W. et al. An analysis of thermal and solar zone radiation models using na Angstrom–Prescott equation and artificial neural networks. **Energy**, v. 33, n. 7, p. 115-1127, 2008.

WEISS, A., HAYS, C. J. Simulation of daily solar irradiance. **Agricultural and Forest Meteorology**, 2004. https://doi.org/10.1016/j.agrformet.2003.12.002

WINSLOW, J. C.; HUNT, E. R. Jr.; PIPER, S. C. A globally applicable model of daily solar irradiance estimated from air temperature and precipitation data. **Ecological Modelling**, v. 143, n. 3, p. 227-243, 2001.

ZANETTI, S. S. et al. Estimação da evapotranspiração de referência no Estado do Rio de Janeiro usando redes neurais artificiais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 12, n. 2, p. 174 – 180, 2008.