

## Desempenho do Modelo Angstrom-Prescott (A-P) e das Técnicas MVS e RNA na Estimativa da Irradiação Solar Global em Botucatu/SP/Brasil

M. B. P. da Silva<sup>1</sup>, J. F. Escobedo<sup>2</sup>, T. J. Rossi<sup>3</sup>, C. M. dos Santos<sup>4</sup>, S. M. G. da Silva<sup>5</sup>

<sup>1,2,3,4</sup>Departamento de Engenharia Rural – FCA/UNESP – Botucatu/SP/Brasil

<sup>5</sup>UNESP Campus Experimental de Registro/SP/Brasil

Tel. + 55143811-7585 e-mail: [mauricio.prado19@hotmail.com](mailto:mauricio.prado19@hotmail.com)

**RESUMO.** No presente trabalho é descrito o estudo comparativo entre métodos de estimativas da irradiação solar global ( $H_G$ ) diário: modelo de Angstrom-Prescott (A-P) e duas técnicas de Aprendizado de Máquina (AM) - Máquinas Vetores de Suporte (MVS) e Redes Neurais Artificiais (RNA). A base de dados de  $H_G$  foi medida no período de 1996 a 2011 em Botucatu/SP/Brazil. A comparação dos indicadores estatísticos MBE, RMSE,  $d$  de Willmott,  $r$  e  $R^2$  obtidos na validação entre os modelos (A-P), MVS e RNA mostrou que: 1- a técnica MVS apresentou melhor desempenho que o modelo estatístico de (A-P) e a técnica RNA; 2- o modelo estatístico (A-P) no geral apresentou melhor desempenho que a técnica RNA.

**Palavras - chave:** Radiação solar, Angstrom-Prescott e modelagem estatística.

### INTRODUÇÃO

O Brasil, país com dimensões continentais, possui enorme escassez da disponibilidade de informações solarimétricas. Por isso, diversos pesquisadores vem desenvolvendo modelos que permitem estimar os diversos tipos de radiação solar por meio de variáveis mais facilmente monitorada nas estações meteorológicas. Em geral, esses modelos são divididos em classes como: estatísticos, físicos (transferência radiativa) e mais recentemente por técnicas de Aprendizado de Máquina (AM) (Escobedo et al., 2012; Oliveira et al., 2002; Soares et al., 2004; Santos et al., 2014; Lyra et al., 2015). Os modelos estatísticos, normalmente são mais recomendados em função da simplicidade e precisão. A desvantagem é que esses modelos tem validade apenas para os locais em que são gerados ou para regiões de clima similar. O principal modelo estatístico é o de Angstrom-Prescott (A-P), que estima a radiação global ( $H_G$ ) a partir do brilho solar ( $n$ ). Os modelos físicos são mais complexos que os modelos estatísticos e exigem medidas com precisão de variáveis meteorológicas na superfície e na atmosfera. Os modelos AM com auxílio da tecnologia computacional, são capazes de solucionar problemas complexos, e têm sido aplicadas com sucesso para previsão de irradiação solar e aplicações solares. A técnica permite, sem a necessidade de conhecer o objeto investigado, modelar um sistema conhecendo apenas os valores de entrada e saída. A maior vantagem do uso de modelos AM é a sua simplicidade e otimização do tempo (Oliveira et al., 2006; Soares e Silva, 2011). As principais técnicas de Aprendizado de Máquina (AM) que vem sendo utilizadas são: Máquina Vetor de Suporte (MVS), Programa Genético (PG), Programa de Expressão do Gene (PEG), Redes Neurais Artificiais (RNA), Sistema de Inferência Neuro-Fuzzy (SINF) e algumas versões híbridas dessas mesmas técnicas (MVS e SINF). Estimativas de  $H_G$ , por meio dos modelos de AM, têm sido estudados

---

<sup>1</sup> Engenheiro Agrônomo, Doutorando em Irrigação e Drenagem, FCA/UNESP/Botucatu.

<sup>2</sup> Físico, Prof. Dr. Departamento de Engenharia Rural, FCA/UNESP/Botucatu.

<sup>3</sup> Graduando em biologia, Departamento de Engenharia Rural, FCA/UNESP/Botucatu.

<sup>4</sup> Físico, Doutor em Irrigação e Drenagem, FCA/UNESP/Botucatu.

<sup>5</sup> Técnico em Processamento de Dados, Doutora em Energia na Agricultura, Professor assistente, Unidade de Registro – UNESP.

comparados com modelos estatísticos e os resultados tem mostrado que o desempenho dos modelos de AM são similares em alguns casos e superiores em outros aos modelos estatísticos (Elizondo et al. 1994; Tymvios et al., 2005; Chen et al., 2013; Piri et al., 2015). No presente trabalho, uma base de dados de irradiação solar global ( $H_G$ ) de 16 anos, período 1996 a 2011, foi utilizada para a obtenção da equação de A-P e para treinamento de duas técnicas de AM (MVS e RNA). Para as técnicas MVS e RNA foram estudados quatro diferentes arquiteturas combinando variáveis meteorológicas de entrada no modelo: o modelo 1 da MVS e RNA usam como variáveis de entrada a razão de insolação ( $n$ ) e  $H_0$  igual ao modelo de A-P; e nos modelos seguintes 2, 3 e 4 foram acrescentadas outras variáveis, uma a uma, como temperatura do ar, precipitação e umidade relativa do ar. Os resultados da validação, comparação entre medidas e estimativas dos modelos por meio de correlação ( $r$ ) e indicativos estatísticos (MBE, RMSE e  $d$  de Willmott) foram apresentados e discutidos.

## METODOLOGIA

### *Instrumentação e medidas*

Os dados de irradiação solar global utilizados neste trabalho foram medidas no período 1996-2011 na Estação de Radiometria Solar na Faculdade de Ciências Agrárias (FCA) da Universidade Estadual Paulista (UNESP). A irradiância solar global ( $I_G$ ,  $W m^{-2}$ ), foi monitorada por um piranômetro Eppley PSP com erro 4,1% (Reda et al., 2008); brilho solar ( $n$ , horas), obtido por um heliógrafo *Campbell-Stokes*; precipitação ( $P$ , mm) medidos através de um pluviógrafo da marca *Ota Keiki Seisakusho*; temperaturas máxima e mínima ( $T$ , °C) medidos por termômetro de bulbo de mercúrio e álcool; e umidade relativa do ar (% UR) medidos através de um higrômetro (seguindo as recomendações sugeridas pela World Meteorological Organization - WMO (1981).

Na aquisição dos dados de  $H_G$  foi utilizado um datalogger CR 23X da Campbell Scientific operando na frequência de 1 Hz e armazenando médias a cada 5 minutos. Estes dados passaram por um rigoroso controle de qualidade (para eliminação de valores espúrios ou inconsistentes) através de programas desenvolvidos para cálculo das irradiações integrados no dia (Chaves e Escobedo, 2000).

### *Aplicação da Máquina Vetor de Suporte (MVS) e Rede Neural Artificial (RNA)*

A MVS é uma técnica interessante pela precisão na estimativa. Existem diferentes algoritmos de aprendizado de máquinas (AM), neste trabalho o *Sequential Minimal Optimization* ( $SMO_{reg}$ ) é utilizado para regressão.  $SMO_{reg}$  é um algoritmo simples que rapidamente soluciona e resolve o problema de menor otimização possível com dois multiplicadores de Lagrange (Platt, 1998; Smola e Schölkopf, 1998). A flexibilidade do SMO na regressão é atribuída ao uso do conjunto de funções de *Kernel*. A escolha adequada da função de *Kernel* é determinante para os dados modelados e são essencialmente funções de similaridade com determinadas propriedades matemáticas. Pela eficiência computacional, simplicidade e adaptação para otimização de problemas mais complexos, neste estudo a *Radial Basis Function* (RBF) de *Kernel* é utilizada para regressão. A RBF é utilizada e recomendada pela simplicidade, quando comparada, por exemplo, com a função polinomial (Dong et al., 2005).

Na aplicação da RBF é necessário ajustar adequadamente os parâmetros  $C$  (custo),  $\gamma$  (gamma) e  $\epsilon$  (épsilon). Os parâmetros  $C$  e  $\gamma$  são mutuamente dependentes entre si, baixos valores de  $C$  produz um aprendizado de máquina com má aproximação e, valores muito altos, geram um aprendizado de máquina mais complexo (Raghavendra e Deka, 2014). Os valores de  $C$  e  $\gamma$  são testados e aquele com a melhor precisão na validação-cruzada são escolhidos. Os modelos gerados com o algoritmo SMO neste trabalho foram com base na função de radial basis function (RBF). Foram testados empiricamente diferentes combinações para os parâmetros da RBF, depois de inúmeros testes os parâmetros utilizados da RBF foram ( $C = 100$ ,  $\gamma = 0,3$  e  $\epsilon = 0,001$ ), valores encontrados na literatura (Ramedani et al., 2014; Mohammadi et al., 2015).

A RNA utilizada foi a *Multilayer perceptron* (MLP). Essa rede é amplamente utilizado em modelagem para resolução de problemas complexos e consiste de um conjunto de unidades sensoriais (nós de fonte) com três ou mais camadas interligadas. Na Figura 1 é demonstrada a estrutura de uma MLP, onde a primeira camada é a entrada ( $x_{i,j}$ ), a segunda possui uma ou mais camadas ocultas de nós

computacionais com pesos de ligações ( $w_{i,j}$ ) e a terceira camada corresponde à saída de nós computacionais ( $y_i$ ), (Zanetti et al., 2008; Lyra et al., 2015).

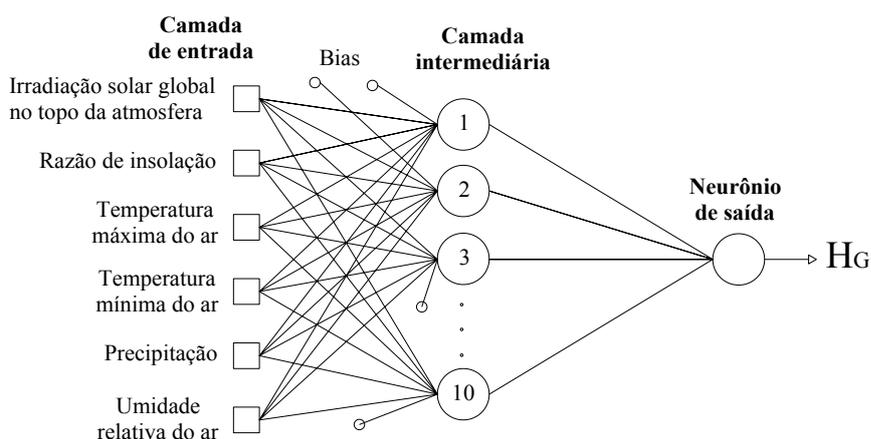


Figura 1. Diagrama em blocos demonstrando a arquitetura da Rede Neural Artificial (RNA) utilizada. Adaptado de Haykin (1998) e Zanetti et al. (2008).

A saída da rede ( $y_i$ ) pode ser modelada conforme Equação (1) (Lam et al., 2008):

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

onde  $x_{i,j}$  é o sinal de entrada a partir de neurônios de ordem  $j$  (para a camada de entrada),  $w_{i,j}$  é o peso da ligação direta do neurônio  $j$  para neurônio  $i$  (na camada oculta) e  $\theta_i$  é a bias do neurônio  $i$ . Na MLP o sinal de entrada se propaga para frente através da rede, camada por camada. Dentro de cada camada há vários neurônios (nós), que estão ligados aos nodos na camada adjacente com pesos diferentes. Cada entrada é multiplicada por um peso de ligação. Após o treinamento os pesos terão informações importantes, antes eles são aleatórios e não têm nenhum significado. A saída dos neurônios é calculada pela aplicação de uma função de ativação, Equação (2), que é tipicamente sigmóide padrão (Rehman e Mohandes, 2008).

$$f(x) = \frac{1}{(1+\exp(-x))} \quad (2)$$

Existem muitos algoritmos de aprendizado específicos para determinado modelos de redes neurais. Neste trabalho a MLP foi treinada usando o algoritmo de aprendizagem *backpropagation*. Este algoritmo baseia-se na regra de aprendizagem por correção de erros e consiste em duas passagens pelas diferentes camadas das redes: um passe para frente, a propagação e um passe para trás, a retropropagação. No passo para frente, um padrão de atividade é aplicado aos nós sensoriais da rede e os seus efeitos se propagam através da rede camada por camada (Rumelhart et al., 1986).

A MVS e a RNA foram avaliadas através da WEKA (*Waikato Environment for Knowledge Analysis*), que consiste num conjunto de algoritmos de AM, disponível em <http://www.cs.waikato.ac.nz/~ml/weka/>. A WEKA contém ferramentas para pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização (Witten et al., 2011). Nos modelos de RNA gerados neste trabalho foram considerados os valores: taxa de aprendizado = 0,3; momentum = 0,2 e numero de interações = 500. As camadas ocultas (*hiddenlayers*) foram testadas variando de 1 – 10, porém o valor padrão do WEKA foi adotado pelo melhor ajuste encontrado. Na WEKA o padrão das camadas ocultas é definido como “a” = [(variáveis de entrada + classes)/2]. As técnicas de AM foram propostas para estimar a irradiação solar global ( $H_G$ ). Os modelos gerados com estas duas técnicas foram comparados entre si.

### Validação e avaliação dos modelos

Na validação dos modelos estatísticos gerados neste trabalho foi utilizada uma base dados de 2 anos, denominadas de ano típico e atípico, obtidos a partir da base de dados total de 16 anos. A seleção dos anos típico e atípico, foi efetuada através de análises estatísticas onde se comparou para cada mês do

ano, o valor da irradiação solar global ( $H_G$ ) média inter-anos e o valor da irradiação média por ano. Os resultados encontrados para todos os meses estão representados na Tabela 1, a qual mostra a constituição dos anos típico e atípico.

		Meses											
		Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
Ano	Típico	2004	2000	2008	2005	2005	2005	2002	2003	2000	2003	2003	2005
	Atípico	2001	2005	2011	2008	2003	1997	2009	1998	2009	2001	1998	2011

Tabela 1. Ano típico e atípico obtido do banco de dados de 16 anos.

Os indicativos estatísticos *Mean Bias Error (MBE)*, *Relative, Root Mean Square Error (RMSE)*, Coeficiente de correlação ( $r$ ), Coeficiente de determinação ( $R^2$ ), e o Índice de Concordância de Willmott ( $d$ ) [Willmott, 1981], foram utilizados para a avaliação da performance dos modelos. Alguns autores em seus trabalhos destacam uma escala classificatória para os diferentes intervalos de rRMSE para avaliar a acurácia dos modelos (Jamieson et al., 1991; Heinemann et al., 2012; Li et al., 2013): Excelente se  $RMSE < 10\%$ ; Bom se  $10\% \leq RMSE < 20\%$ ; Aceitável se  $20\% \leq RMSE < 30\%$ ; Pobre se  $RMSE \geq 30$ .

## RESULTADOS E DISCUSSÃO

### *Correlação entre a transmissividade atmosférica da irradiação global ( $H_G/H_0$ ) e a razão de insolação ( $n/N$ ) diária: equação de Angstrom-Prescott (A-P) e treinamento da RNA e MVS.*

A Figura (2) mostra a correlação entre a transmissividade atmosférica da irradiação global ( $H_G/H_0$ ) e a razão de insolação ( $n/N$ ) diária, ( $H_G/H_0$ ) x ( $n/N$ ), para a base de dados medida no período de 1996-2011 em Botucatu, SP. A correlação é linear em todo intervalo de variação de  $n/N$  entre 0 e 1. O espalhamento da correlação é similar a maioria dos trabalhos da equação de Angstrom-Prescott no mundo (Martinez-Lozano et al., 1984).

O espalhamento na correlação é elevado, onde para cada valor de ( $n/N$ ) há um largo intervalo de variação nos valores ( $H_G/H_0$ ). O efeito é devido a grande variabilidade das concentrações de nuvens, vapor de água e aerossóis na atmosfera, que absorvem e espalham a radiação global, diferentemente, nos diversos tipos de cobertura de céu para mesmos valores de  $n/N$ .

A Equação de (A-P) obtida na correlação da Figura (2), por meio de regressão linear é dada por:

$$(H_G/H_0) = 0,253 + 0,465(n/N) \quad (3)$$

O coeficiente de determinação  $R^2 = 0,806$ , bem como da transmissividade atmosférica mínima ( $H_G/H_0$ ) = 0,253 ou a transmissividade atmosférica máxima ( $H_G/H_0$ ) = 0,718 é similar aos valores obtidos em mais de 100 localidades do Brasil, e estão dentro do intervalo de variação de  $a$  ( $0,26 \pm 0,053$ ),  $a+b$  ( $0,70 \pm 0,07$ ) e  $R^2$  cuja média é de:  $0,78 \pm 0,10$  (Silva, 2016).

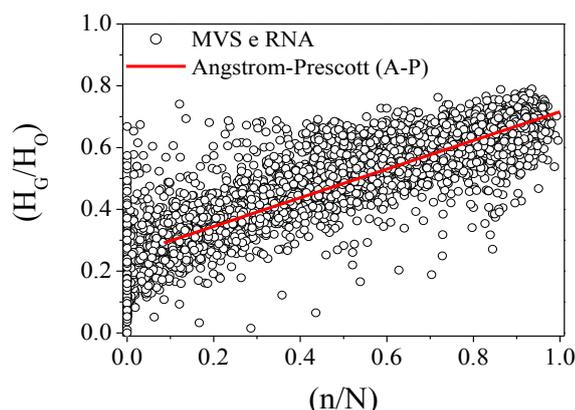


Figura 2. Correlação entre a transmissividade atmosférica da irradiação solar global ( $H_G/H_0$ ) e a razão de insolação ( $n/N$ ), e a reta obtida por regressão linear no período de 1996-2011.

Fatores climáticos, como tipo e espessuras de nuvens, concentrações de vapor de água e aerossóis na atmosfera são os fatores responsáveis pelas variações de  $a$  e  $a+b$  em cada local. O resultado encontrado neste trabalho é próximo ao sugerido por Allen et al. (1998) através do Boletim FAO-56 que é  $a = 0,25$  e  $b = 0,50$  para ser uma equação universal que atende locais onde não existem medidas da radiação global.

Modelos		Variáveis meteorológicas de entrada
MVS-RBF	RNA-MLP	
MVS1	RNA1	$H_0, n/N$
MVS2	RNA2	$H_0, T_{\text{máx}}, T_{\text{mín}}, n/N$
MVS3	RNA3	$H_0, T_{\text{máx}}, T_{\text{mín}}, P, n/N$
MVS4	RNA4	$H_0, UR, T_{\text{máx}}, T_{\text{mín}}, P, n/N$

Tabela 2. Arquitetura dos Modelos gerados pelas técnicas Máquina Vetor de Suporte (MVS) e Redes Neurais Artificiais (RNA) com diferentes variáveis meteorológicas de entrada.

A correlação (Figura 2) possibilitou o treinamento das técnicas MVS e RNA respectivamente. O modelo de Angstrom-Prescott e as técnicas MVS1 e RNA1 (modelo I) utilizaram as mesmas variáveis de entrada  $H_0, N$  (calculados) e  $n$  (medido). Além do modelo I, foram treinados 3 novos modelos (MVS2 a MVS4 e RNA2 a RNA4), acrescentando na sequência os seguintes parâmetros meteorológicos (Tabela 2) temperaturas máxima e mínima do ar ( $T, ^\circ\text{C}$ ) precipitação ( $P, \text{mm}$ ) e umidade relativa ( $UR, \%$ ), como mostra a Tabela (2).

#### Validação do modelo de Angstrom-Prescott (A-P).

A Figura (3) mostra as correlações obtidas na validação entre estimativas e medidas diárias de  $H_G$ , para as bases de dados anuais típico e atípico. As distribuições dos valores estimados pela equação (A-P) e as medidas nas duas condições de validações estão em concordância linear com as curvas ideais de  $45^\circ$  nos ajustes. As equações de regressão linear com os coeficientes de correlação  $r^t = 0,942$ , e  $r^a = 0,939$ , mostram que a equação (A-P) pode estimar  $H_G$  com coeficientes de determinação  $(R^2)^t = 0,887$  (Figura 3 A) e com  $(R^2)^a = 0,882$  (Figura 3 B). Os coeficientes de correlação linear ( $r$ ) obtido neste trabalho foram superiores ao valor de  $r = 0,87$ , obtido por Nicácio et al. (2001), e são da mesma ordem de grandeza de  $r = 0,89$  a  $0,98$  determinados por Li et al. (2011).

Os indicativos estatísticos MBE, RMSE e  $d$  de Willmott, obtidos na validação foram: para  $MBE^t_{(A-P)} = -3,0\%$  e  $MBE^a_{(A-P)} = 1,1\%$ . O resultado está na ordem de grandeza do erro experimental do equipamento de medida da radiação global, que é de  $4,1\%$  (Reda et al., 2008). Os valores de MBE para Botucatu são superiores aos valores encontrados por Nicácio et al. (2001); Martim et al. (2014); Sabziparvar et al. (2013); Zhao et al. (2013) e Iziomon e Mayer (2001). Por outro lado, os resultados foram inferiores aos valores encontrados por: Beruski et al. (2015); Lam et al. (2008); Wan et al. (2008) e Li et al. (2011). Os resultados obtidos por Manzano et al. (2015), alternam valores superiores e inferiores aos resultados de Botucatu.

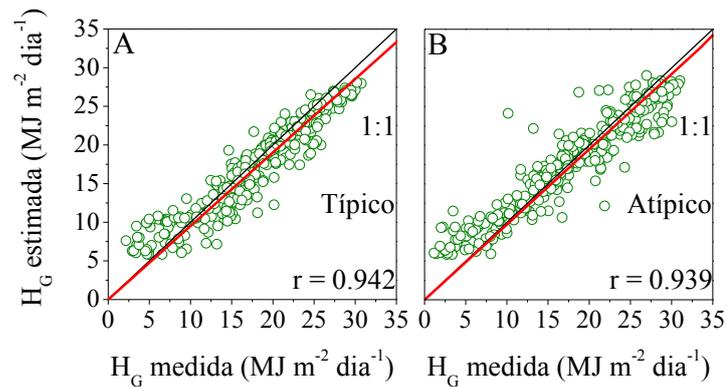


Figura 3. Comparação entre os valores estimados por meio de Angstrom-Prescott (A-P) e medidos de  $H_G$  para anos típico e atípico.

Para o espalhamento  $RMSE^t_{(A-P)} = 13,1\%$  e  $RMSE^a_{(A-P)} = 15,7\%$ . Os valores de RMSE estatisticamente também foram considerados bons resultados segundo os critérios estabelecidos por Jamieson et al. (1991) e Heinemann et al. (2012). Os valores de RMSE local foram superiores aos valores encontrados por Iziomon e Mayer (2001); Nicácio et al. (2001); Li et al. (2011); Torres et al. (2010); Andrade Júnior et al. (2012); Martim et al. (2014); Beruski et al. (2015). Os valores de RMSE local alternam ainda valores aos obtidos por: Lam et al. (2008); Liu et al. (2009); Wan et al. (2008); Li et al. (2012); por Chen et al. (2013); por Zhao et al. (2013); por Mohammadi et al. (2015); Sabziparvar et al. (2013); Manzano et al. (2015) e por Park et al. (2015). Os valores de RMSE local foram inferiores encontrados por Pereira (2010). O índice de concordância  $d^t_{(A-P)} = 0,963$  e  $d^a_{(A-P)} = 0,959$ , próxima de 1 mostram boa concordância entre a estimativa e a medida de  $H_G$ .

#### Validação dos modelos de treinamento das técnicas de AM (MVS e RNA)

A Figura 4 mostra diagrama de dispersão entre os valores estimados de  $H_G$  pelos modelos MVS e RNA e as medidas, bem como as retas obtidas por regressão linear e coeficientes de determinação( $r$ ), para os anos típico( $r^t$ ) e atípico( $r^a$ ) respectivamente.

Os valores estimados por MVS1 e RNA1 estão em concordância linear com as medidas (Figura 4 A, B): as retas de correlações ( $r$ ) estão muito próximas da reta ideal ( $45^\circ$ ) ou  $r = 1$  no ajuste. Para MVS1 e RNA1 os valores de  $r^t = 0,962$  e  $r^t = 0,924$  e  $r^a = 0,947$  e  $r^a = 0,932$ , respectivamente, mostram que a medida de  $H_G$  e a estimativa de  $H_G$  estão estatisticamente bem correlacionadas. A técnica MVS1 teve melhor desempenho que a RNA1 nas duas condições de validações.

No modelo 2, as técnicas MVS2 e RNA2, (Figura 4 C, D) também apresentam correlações ( $r$ ) muito próximas da reta ideal ( $45^\circ$ ) no ajuste, nos quais os valores de  $r^t = 0,966$  e  $r^t = 0,965$  e  $r^a = 0,951$  e  $r^a = 0,951$ , mostram que a medida e a estimativa de  $H_G$  estão muito bem correlacionados. A inclusão das temperaturas ( $T$ ,  $T_{m\acute{a}x}$  e  $T_{m\acute{i}n}$ ) no treinamento das técnicas de AM do modelo1 ( $H_0$  e  $n/N$ ) aumentaram os valores dos coeficientes de correlações para os dois modelos MVS2 e RNA2. A combinação das variáveis  $H_0$ ,  $n/N$ ,  $T_{m\acute{a}x}$  e  $T_{m\acute{i}n}$ , nos modelos MVS2 e RNA2, representam a junção de todas as variáveis das equações de Angstrom-Prescott (A-P), Hargreaves-Samani e Bristow-Campbell. Essas duas últimas equações usam as variáveis  $T_{m\acute{a}x}$  e  $T_{m\acute{i}n}$  na estimativa de  $H_G$  com altos valores de  $r$  (Hargreaves e Samani, 1982; Bristow e Campbell, 1984).

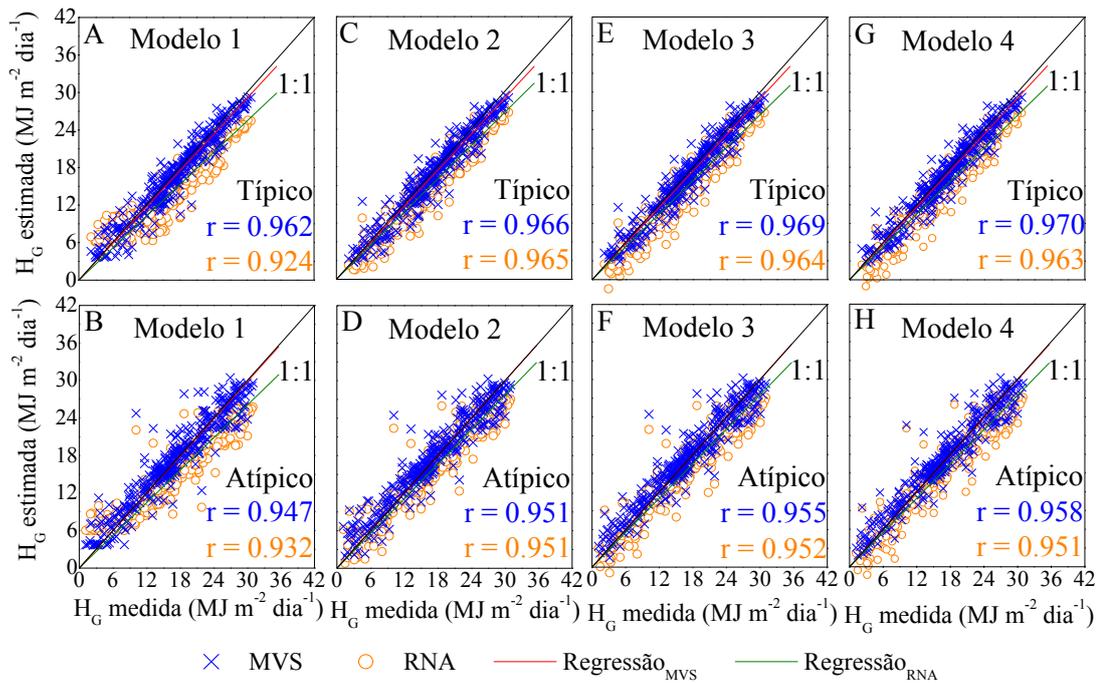


Figura 4. Correlações entre os valores estimados pelos modelos de AM, e os valores medido de  $H_G$ .

Essa combinação foi positiva, porque os valores de  $r$  aumentaram em relação a MVS1 e RNA1. O aumento maior foi com a RNA2, que apresentou uma maior amplitude de variação, o valor de  $r^t = 0,924$  aumentou para  $r^t = 0,965$ , enquanto  $r^a = 0,932$  para  $r^a = 0,951$ . Mesmo assim a comparação nos valores de  $r$  mostra que a MVS2 ainda possui melhor desempenho que a rede RNA2 nas duas condições de validações. Similarmente para o modelo MVS3 e RNA3 (Figura 4 E, F) os valores de  $r^t = 0,969$  e  $r^t = 0,955$ , e de  $r^a = 0,964$  e  $r^a = 0,952$  respectivamente, apresentaram também elevadas correlações, nas duas condições de validação. A introdução da precipitação no modelo 3 aumentaram pouco os valores de  $r$  dos modelos MVS3 e RNA3. Para o modelo MVS4 e RNA4 (Figura 4 G, H) os valores de  $r^t = 0,970$  e  $r^t = 0,958$ , e de  $r^a = 0,963$  e  $r^a = 0,951$  respectivamente, apresentam também elevadas correlações estatísticas nas duas condições de validação. A introdução da umidade relativa (%) no modelo 3 (MVS3 e RNA3) manteve os valores de  $r$ , aproximadamente iguais nas duas condições de validações.

Os indicativos estatísticos MBE, RMSE e  $d$  de Willmott, obtidos da comparação dos valores estimados gerados com as técnicas de AM e as medidas de  $H_G$  respectivamente para os anos típico e atípico, são mostrados na Figura (5):

Para o modelo 1, os valores de MBE (Figura 5 A) para os anos típico e atípico respectivamente mostram que  $MBE^t_{MVS1} = -2,6\%$  e  $MBE^a_{MVS1} = 1,6\%$ , enquanto a  $MBE^t_{RNA1} = -14,1\%$  e  $MBE^a_{RNA1} = -10,7\%$ . Os valores de RMSE (Figura 5 B):  $RMSE^t_{MVS1} = 10,5\%$  e  $RMSE^a_{MVS1} = 13,9\%$ , enquanto para  $RMSE^t_{RNA1} = 20,3\%$  e  $RMSE^a_{MVS1} = 19,9\%$ . O índice de concordância ( $d$ ) de Willmott (Figura 5 C) para MVS1 foi superior em seqüência para  $d^t = 0,98$  e  $d^a = 0,97$ , enquanto que para RNA1  $d^t = 0,91$  e  $d^a = 0,93$  respectivamente.

Para o modelo 2 os valores de MBE (Figura 5 D) mostram que  $MBE^t_{MVS2} = -3,3\%$  e  $MBE^a_{MVS2} = 2,3\%$ , enquanto que  $MBE^t_{RNA2} = -11,6\%$  e  $MBE^a_{RNA2} = -6,3\%$ . Os valores de RMSE (Figura 5 E) foram:  $RMSE^t_{MVS2} = 10,2\%$  e  $RMSE^a_{MVS2} = 13,5\%$ , enquanto que para a técnica  $RMSE^a_{RNA2} = 15,1\%$  e  $MSE^a_{RNA2} = 15,0\%$ . O índice de concordância ( $d$ ) de Willmott (Figura 5 F) com a MVS2 foi superior em seqüência para  $d^t = 0,98$  e  $d^a = 0,97$ , enquanto que para RNA2  $d^t = 0,96$  e  $d^a = 0,97$  respectivamente.

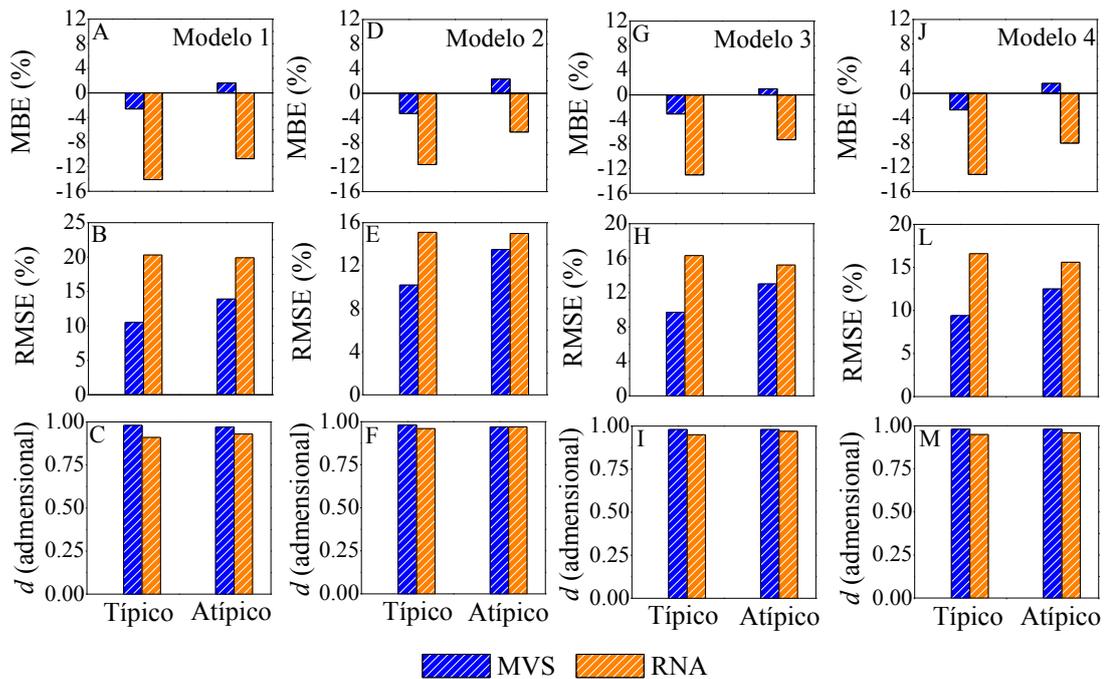


Figura 5. Indicativos estatísticos MBE (%), RMSE (%) e  $d$  de Willmott na validação.

Para o modelo 3 os valores de MBE (Figura 5 G) mostram que  $MBE_{MVS3}^t = -3,1\%$  e  $MBE_{MVS3}^a = 2,2\%$ , enquanto que  $MBE_{RNA3}^t = -13,0\%$  e  $MBE_{RNA3}^a = -7,0\%$ . Os valores de RMSE (Figura 5 H) foram:  $RMSE_{MVS3}^t = 9,7\%$  e  $RMSE_{MVS3}^a = 13,0\%$  enquanto para  $RMSE_{RNA3}^t = 16,3\%$  e  $RMSE_{RNA3}^a = 15,6\%$ . O índice de concordância ( $d$ ) de Willmott (Figura 5 I) com a MVS3 foi igual para  $d^t = 0,98$  e  $d^a = 0,98$ , enquanto que para RNA1  $d^t = 0,95$  e  $d^a = 0,97$  respectivamente.

Para o modelo 4 os valores de MBE (Figura 5 J) mostram que  $MBE_{MVS4}^t = -2,7\%$  e  $MBE_{MVS4}^a = 1,6\%$ , enquanto que  $MBE_{RNA4}^t = -13,2\%$  e  $MBE_{RNA4}^a = -8,1\%$ . Os valores de RMSE (Figura 5 L) foram:  $RMSE_{MVS4}^t = 9,5\%$  e  $RMSE_{MVS4}^a = 12,5\%$ , enquanto que  $RMSE_{RNA4}^t = 16,6\%$  e  $RMSE_{RNA4}^a = 15,6\%$ . O índice de concordância ( $d$ ) de Willmott (Figura 5 M) com a MVS4 foi igual para  $d^t = 0,98$  e  $d^a = 0,98$ , enquanto que para RNA1  $d^t = 0,95$  e  $d^a = 0,96$  respectivamente.

#### Comparação do desempenho entre os modelos Angstrom-Prescott (A-P) e as técnicas MVS e RNA em Botucatu e outros locais.

Os valores dos indicativos estatísticos MBE, RMSE,  $d$  de Willmott,  $r$  e  $R^2$  do modelo de (A-P) e das técnicas MVS e RNA, obtidas em Botucatu, mostram que o modelo MVS4 possui melhor desempenho que o modelo estatístico de (A-P) e o modelo RNA4 nas duas condições de validação, respectivamente (Figura 6). A comparação entre MVS4 x (A-P) mostra que: os valores de  $MBE_{MVS4}^t = -2,7\%$  e  $MBE_{MVS4}^a = 1,6\%$  são inferiores a  $MBE_{(A-P)}^t = -3,0\%$  e  $MBE_{(A-P)}^a = 1,1\%$  (Figura 6 A); os valores de  $RMSE_{MVS4}^t = 9,4\%$  e  $RMSE_{MVS4}^a = 12,5\%$  são inferiores aos de  $RMSE_{(A-P)}^t = 13,1\%$  e  $RMSE_{(A-P)}^a = 15,7\%$  (Figura 6 B); os valores de  $d$  de Willmott  $d_{MVS4}^t = 98,0\%$  e  $d_{MVS4}^a = 98,0\%$  são superiores aos de  $d_{(A-P)}^t = 95,0\%$  e  $d_{(A-P)}^a = 96,0\%$  (Figura 6 C); os valores de  $r_{MVS4}^t = 0,970$  e  $r_{MVS4}^a = 0,918$  são superiores aos de  $r_{(A-P)}^t = 0,942$  e  $r_{(A-P)}^a = 0,939$  e os valores de  $(R^2)_{MVS4}^t = 0,941$  e  $(R^2)_{MVS4}^a = 0,918$  são superiores aos de  $(R^2)_{(A-P)}^t = 0,887$  e  $(R^2)_{(A-P)}^a = 0,882$ .

Os resultados obtidos neste trabalho estão de acordo com os resultados dos trabalhos por Chen et al. (2013) na China, Mohammadi et al. (2015) no Irã e por Piri et al. (2015) no Irã. Em todos trabalhos citados a MVS possui melhor performance que o modelo (A-P). O valor de RMSE para Botucatu foram inferiores aos resultados obtidos por Chen et al. (2013) na China. E são da mesma ordem de grandeza aos resultados obtidos por Mohammadi et al. (2015) no Irã e por Piri et al. (2015) no Irã.

Os valores dos indicativos estatísticos MBE, RMSE,  $d$  de Willmott,  $r$  e  $R^2$  na comparação entre (A-P) x RNA4 para os anos típico e atípico, mostram que o modelo (A-P) apresentam no geral melhor desempenho que o modelo RNA4.

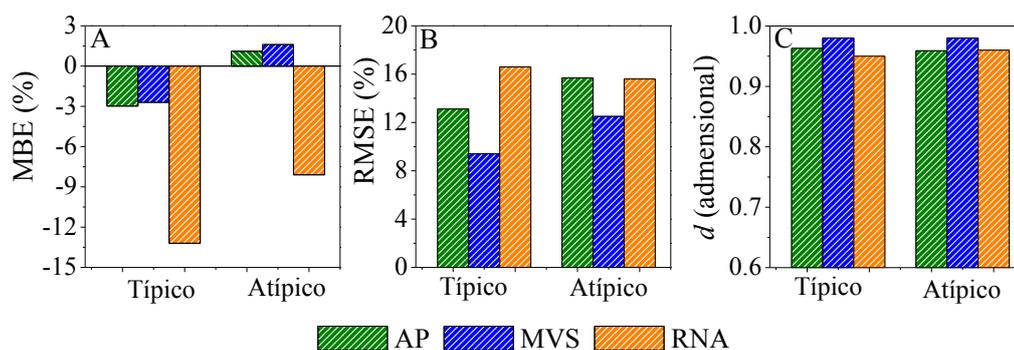


Figura 6. Indicativos estatísticos MBE, RMSE e  $d$  de Willmott para anos típico e atípico.

A comparação entre os modelos (A-P) x RNA4 mostra que os valores de  $MBE_{(A-P)}^t = -2,98,0\%$  e  $MBE_{(A-P)}^a = 1,1\%$  são inferiores aos valores  $MBE_{RNA4}^t = -13,2\%$  e  $MBE_{RNA4}^a = -8,1\%$  (Figura 6 A); os valores de  $RMSE_{(A-P)}^t = 13,1\%$  e  $RMSE_{(A-P)}^a = 15,7\%$ , são inferiores a  $RMSE_{RNA4}^t = 16,6\%$  e  $RMSE_{RNA4}^a = 15,6\%$ , os valores de  $d$  de Willmott  $d_{(A-P)}^t = 95,0\%$  e  $d_{(A-P)}^a = 96,0\%$  (Figura 6 B); é igual a  $d_{RNA4}^t = 95,0\%$  e  $d_{RNA4}^a = 96,0\%$  (Figura 6 C); os valores de  $r_{(A-P)}^t = 0,942$  e  $r_{(A-P)}^a = 0,939$  são inferiores a  $r_{RNA4}^t = 0,963$  e  $r_{RNA4}^a = 0,951$  e para o coeficiente de determinação  $(R^2)_{(A-P)}^t$  igual a  $0,887$  e  $(R^2)_{(A-P)}^a = 0,882$  são inferiores a aos valores  $(R^2)_{RNA4}^t = 0,927$  e  $(R^2)_{RNA4}^a = 0,904$ . Os resultados obtidos para os indicativos obtidos neste trabalho estão de acordo e nas mesma ordem de grandeza com os resultados do trabalho de de Wan et al. (2008) para 41 cidades na China.

## CONCLUSÃO

Dos resultados apresentados e discutidos neste trabalho conclui-se que: a equação de estimativa de  $H_G$  de Angstrom-Prescott (A-P), para Botucatu, SP, com coeficientes de determinação  $R^2 = 0,806$  é similar aos resultados alcançados em outras localidades. Os indicativos estatísticos: MBE inferior a  $3,0\%$ ; RMSE inferior a  $15,68\%$ ;  $d$  de Willmott superior a  $0,95$ , indicam que o modelo (A-P) pode ser utilizado na estimativa de  $H_G$  com precisão e exatidão nas duas condições de validação;

As técnicas MVS1 e RNA1 (modelo 1), com as mesmas variáveis de entrada  $H_0$  e  $(n/N)$  que o modelo (A-P), mostraram, que podem estimar  $H_G$  na mesma ordem de grandeza que a equação de Angstrom-Prescott clássica;

Os modelos MVS2 e RNA2 tiveram aumentos nos valores de  $r$  com a inserção das temperaturas ( $T$ ,  $T_{máx}$  e  $T_{mín}$ ) nos modelos MVS1 e RNA1. Os modelos MVS3 e RNA3, bem como MVS4 e RNA4, tiveram melhoras no desempenho com as inclusões da precipitação e umidade relativas nos modelos 3 e 4 respectivamente;

A comparação entre os modelos (A-P), MVS4 e RNA4 por meio dos valores dos indicativos estatísticos MBE, RMSE,  $d$  de Willmott,  $r$  e  $R^2$  obtidos da validação mostrou que: entre o modelo (A-P) e MVS4, a técnica MVS4 apresentou melhor desempenho que o modelo estatístico de (A-P); entre os modelos MVS4 e RNA4, a técnica MVS4 apresentou melhor desempenho que a técnica RNA4; entre o modelo (A-P) e a RNA4, o modelo estatístico (A-P) apresentou no geral melhor desempenho que RNA4.

## REFERÊNCIAS

- Allen, R. G. et al. (1998). Crop evapotranspiration - guidelines for computing crop water requirements. pp. 300. Food and Agriculture Organization of the United Nations (FAO) - Irrigation and drainage paper 56, Rome.
- Alvarez, R. (2009). Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. European Journal Agronomy, v. 30, 70-77.
- Andrade Júnior et al. (2012). Coeficientes da equação de Angström-Prescott para Parnaíba, Piauí. Comunicata Scientiae, n. 3, v. 1, 50-54.
- Beruský et al. (2015). Desempenho de diferentes modelos de estimativa da radiação solar global em Ponta Grossa, PR. Revista Brasileira de Meteorologia, v. 30, n. 2, 83-91.

- Bristow, K. L. e Campbell, G. S. (1984). On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*, v. 31, n. 2, 159-166.
- Chaves, M. A. e Escobedo, J. F. (2000). A software to process daily solar radiation data. *Renewable Energy*, v. 19, 339 - 344.
- Chen, J. L.; LI, G. S.; Wu, S. J. (2013). Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. *Energy Conversion and Management*, v. 75, 311-318.
- Dong, J. X.; Devroye, L.; Suen, C.Y. (2005). Fast SVM training algorithm with decomposition on very large data sets. *Pattern Analysis and Machine Intelligence*, v. 27, n.4, 603-618.
- J. F. Escobedo, E. T. Teramoto; A. P. Oliveira, J. Soares, G. Codato: Equações de estimativa das frações solar direta (kdh) e difusa (kd) em função do índice de claridade(kt) e razão de insolação (n/n), *Avances en Energías Renovables y Medio Ambiente*, Vol. 16, 2012.
- Elizondo, D.; Hoogenboom, G.; Mcclendon, R. W. (1994). Development of a neural network model to predict daily solar radiation. *Agricultural and Forest Meteorology*, v. 71, 115-132.
- Hargreaves, G. H. e Samani, Z. A. (1982). Estimating potential evapotranspiration. *Journal of Irrigation and Drainage Engineering*, v. 108, 225-230.
- Haykin, S. (1998). *Neural networks: A comprehensive foundation*. 2nd ed. pp. 897. Hamilton: Prentice Hall.
- Heinemann et al. (2012). Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. *Bragantia, Campinas*, v. 71, n. 4, 572-582.
- Iziomon, M. G.e Mayer, H. (2001). Performance of solar radiation models - a case study. *Agricultural and forest meteorology*, v. 110, 1-11.
- Jamieson, P. D.; Porter, J. R.; Wilson, D. R. (1991). A test of the computer simulation model ARC - WHEAT1 on wheat crops grown in New Zealand. *Field Crops Research*, v.27, 337-350.
- Kumar, R. (2015). Comparison of regression and artificial neural network models for estimation of global solar radiations. *Renewable and Sustainable Energy Reviews*, v. 52, 1294-1299.
- Lam et al. (2008). Climatic influences on solar modeling in China. *Renewable Energy*, v. 33, 1591-1604.
- Li et al. (2011). Global solar radiation estimation with sunshine duration in Tibet, China. *Renewable Energy*, v. 36, 3141-3145.
- Li et al. (2012). Impacto of time interval on the Angstrom-PreScott coefficients and their interchangeability in estimating radiation. *Renewable Energy*, v. 44, 431 - 438.
- Li et al. (2013). General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Conversion and Management*, v. 70, 139 - 148.
- Liu et al. (2009). Calibration of the Angström-PreScott coefficients ( $a$ ,  $b$ ) under different times scales and their impacts in estimating global solar radiation in the Yellow River Basin. *Agricultural Forest and Meteorology*, v. 149, 697-710.
- Lyra et al. (2015). Estimation of monthly global solar irradiation using the Hargreaves-Samani model and na artificial neural network for the state of Alagoas in northeastern Brazil. *Theoretical Applied Climatology*, DOI 10.1007/s00704-015-1541-8.
- Manzano et al. (2015). A single method to estimate the daily global solar radiation from monthly data. *Atmospheric Research*, v. 166, 70 - 82.
- Martim et al. (2014). Coeficientes de Angstrom-PreScott para a Região Norte do Mato Grosso: primeira aproximação. In: XLII Congresso Brasileiro de Engenharia Agrícola – Campo Grande.
- Martínez-Lozano et al. (1984). The historical evolution of the Angstrom Formula and its modifications. *Agricultural and Forest Meteorology*, v.33, n.2-3, 109 - 128.
- Mohammadi et al. (2015). Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion and Management*, v. 91, 433 - 441.
- Nicácio, R. M.; Souza, J. L.; Bernardo, S. O. (2001). Estimativa da irradiância solar global para Maceió utilizando o modelo linear de Angstrom-PreScott. In: XII Congresso Brasileiro de Agrometeorologia - Fortaleza.
- Oliveira, A. P.; Escobedo, J. F.; Machado, A. J.; Soares, J. (2002). Correlation model of diffuse solar-radiation applied to the city of São Paulo, Brazil. *Applied Energy*, v. 71, 59-73.

- Oliveira, A. P.; Soares, J.; Boznar, M. Z.; Mlakar, P.; Escobedo, J. F. (2006). An application of Neural Network technique to correct the dome temperature effects on pyrgeometer measurements. *Journal of Atmospheric and Oceanic Technology*, v. 23, 80-89.
- Park, J.; Das, A.; Park, J. H. (2015). A new approach to estimate the spatial distribution of solar radiation using topographic factor and sunshine duration in South Korea. *Energy Conversion and Management*, v. 101, 30-39.
- Piri et al. (2015). Prediction of the solar radiation on the Earth using support vector regression technique. *Infrared Physics & Technology*, v. 68, 179-185.
- Pereira et al. (2010). Estimativa da radiação solar global para a Região de Pedra Azul. *Enciclopédia Biosfera*, v. 6, n. 11, 1 - 9.
- Platt, J. (1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. Redmond: Microsoft Research, TechReport. pp. 21.
- Ramedani et al. (2014). Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews*, v. 39, 1005-1011.
- Raghavendra, S. e Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, v.19, 372 - 386.
- Reda, I.M.; Myers, D.R.; Stoffel, T.L. (2008). Uncertainty estimate for the outdoor calibration of solar pyranometers: NCSLI Measure. *The Journal of Measurement Science*, v. 3, n. 4, 58-66.
- Rehman, S. e Mohandes, M. (2008). Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy*, v. 36, 571-576.
- Rumelhart, D. E.; Hinton, G. E., Williams, R. J. (1986). Learning representations by backpropagation errors. *Nature*, London, UK, v. 323, 533 - 536.
- Sabzipavar et al. (2013). An improved estimation of the Angstrom-Prescott radiation coefficients for the FAO 56 Penman-Monteith evapotranspiration method. *Water Resources Management*, v. 27, 2839-2854.
- Santos, C. M.; Souza, J. L.; Ferreira Júnior, R. A.; Tiba, C.; Melo, R. O.; Lyra, G. B.; Teodoro, I.; Lyra, G.B.; Lemes, M. A. M. (2014). On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. *Energy*, v. 71, 338 - 398.
- Silva, M. B. P. Estimativa da irradiação solar global pelo método de Angstrom-Prescott e Técnicas de Aprendizado de Máquinas. 2016. pp. 72. Dissertação (Mestrado em Agronomia/Irrigação e Drenagem) - Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu.
- Soares, P. L. B.; Silva, J. P. (2011). 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, 58-72.
- Soares, J.; Oliveira, A. P.; Boznar, M. Z.; Mlakar, P.; Escobedo, J. F.; Machado, A. J. (2004). Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique, v.79, 201-2014.
- Smola, A. J e Schölkopf, B. (2004). A tutorial on Support Vector Regression. *Statistics and Computing*, v.14, 199 - 222.
- Torres et al. (2010). Determinação dos coeficientes do modelo de Angstrom-Prescott para a Região de Canavieiras, Estado da Bahia. *Enciclopédia Biosfera*, v. 6, n. 11, 1 – 7.
- Tymvios et al. (2005). Comparative study of Angstrom and artificial neural networks' methodologies in estimating global solar radiation. *Solar Energy*, v. 78, 752 – 762.
- Wan et al. (2008). An analysis of thermal and solar zone radiation models using an Angstrom-Prescott equation and artificial neural networks. *Energy*, v. 33, 1115-1127.
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, v.2, 184 - 194.
- Witten, I. H.; Frank, E.; Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. 3 rd ed. pp. 630.
- World Meteorological Organization. (1981). *Meteorological Aspects of the utilization of Solar Radiation as an Energy Source*. World Meteorological Organization Technical Note No. 172, WMO-No. 557, Geneva, pp. 298.
- Zanetti et al. (2008). Estimativa da evapotranspiração de referência no Estado do Rio de Janeiro usando redes neurais artificiais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 12, 174-180.

Zhao, N., Zeng, X., Han, S. (2013). Solar radiation estimation using sunshine hour and air pollution index in China. *Energy Conversion and Management*, v. 76, 846-851.

## **Performance of the Angstrom-Prescott (AP) Model and SVM and ANN Techniques to Estimate Global Solar Irradiation in Botucatu/SP/Brazil**

**ABSTRACT.** This paper describes the comparative study of methods to estimate daily ( $H_G$ ) global solar irradiation ( $H_G$ ): Angstrom-Prescott (AP) model and two machine learning techniques (ML) - Support Vector Machines (SVM) and Artificial Neural Networks (ANN). Database was measured from 1996 to 2011, in Botucatu/SP/Brazil. The comparison of MBE, RMSE,  $d$  Willmott,  $r$  and  $R^2$  statistical indicators obtained from the validation of the models (A-P), SVM and ANN showed that: the SVM technique showed better results than the (A-P) statistical model and ANN technique; overall, the (A-P) statistical model showed better performance than the ANN technique.

**Key words:** Solar radiation; Angstrom-Prescott; Statistical Modeling.