

MODELOS DE ESTIMATIVA DA IRRADIAÇÃO SOLAR GLOBAL MENSAL: ANGSTROM-PRESCOTT E APRENDIZADO DE MÁQUINAS EM BOTUCATU/SP/BRASIL

M. B. P. da Silva¹, J. F. Escobedo², A. S. Marques³, T. J. Rossi⁴, C. M. dos Santos⁵

^{1,2,4}Departamento de Engenharia Rural – FCA/UNESP – Botucatu/SP/Brasil

³Instituto Federal de São Paulo - IFSP/Birigui/SP/Brasil

⁵FEA/UFPA/Altamira/PA/Brasil

Tel. + 55 (14) 3811-7585 e-mail: mauricio.prado19@hotmail.com

Recibido 26/07/17, aceptado 25/09/17

RESUMO. No presente trabalho é descrito o estudo comparativo, entre o métodos estatístico de Angstrom-Prescott (A-P)^m e duas técnicas de Aprendizado de Máquina (AM) - Support Vector Machine (SVM)^m e Artificial Neural Network (ANN)^m, nas estimativa da irradiação solar global (H_G) mensal. A base de dados de H_G usada na modelagem foi medida no período de 1996 a 2011 em Botucatu/SP/Brasil. A equação (A-P)^m obtida $(H_G/H_0) = 0,311 + 0,366 * (n/N)$ com $R^2 = 0,710$ foi validada usando duas bases de dados: típica e atípica. Os indicativos estatísticos r e $rRMSE$, obtidos na comparação entre a estimativa e medida foram: $r^t=r^a=0,980$, $rRMSE^t=3,9\%$ e $rRMSE^a=9,8\%$. As técnicas SVM^m e ANN^m foram treinadas em quatro combinações de entradas de variáveis meteorológicas, e validadas nas bases de dados típica e atípica. A comparação dos indicativos estatísticos mostra que: a técnica SVM possui melhor desempenho que o modelo (A-P) e a técnica ANN em estimar H_G . O modelo (A-P) possui melhor desempenho que a técnica ANN em estimar H_G .

Palavras - chave: Radiação solar, Angstrom-Prescott; inteligência artificial.

INTRODUÇÃO

O Brasil, país de dimensão continental, vem aumentando a rede solarimétrica no território com objetivo de conhecer melhor o potencial solar disponível, para estudos nas áreas de conversões da energia solar como térmica, fotovoltaica e biomassa; na agricultura nos modelos de crescimento e produtividade de cultura agrícolas, e estimativas de evapotranspiração (Hsiao et al., 2008; Bosch et al., 2008), entre varias outras. No entanto, em locais onde a radiação global não é monitorada, o uso de modelos de estimativa é a alternativa mais usada nos projetos de pesquisa. Por isso, diversos pesquisadores vêm desenvolvendo e ajustando modelos que permitem estimar a radiação solar global a partir da medida de outra radiação ou de outra variável mais facilmente monitorada nos postos meteorológicos. Em geral, esses modelos são divididos em diferentes classes: estatísticos, físicos (transferência radiativa) e mais recentemente as técnicas de Aprendizado de Máquina (AM) (Oliveira et al., 2002; Soares et al., 2004; Escobedo et al., 2012; Santos et al., 2016).

Os modelos de AM são capazes de solucionar problemas complexos e têm sido aplicados com sucesso para previsão de irradiação solar em aplicações solares. A técnica permite modelar um sistema conhecendo

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

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

³ Engenheiro da Computação, Doutorando em Energia na Agricultura, FCA/UNESP/Botucatu.

⁴ Graduando em biologia, Departamento de Engenharia Rural, FCA/UNESP/Botucatu.

⁵ Físico, Doutor em Irrigação e Drenagem, FCA/UNESP/Botucatu.

apenas as variáveis de entrada e saída. A maior vantagem do uso de modelos de AM é a sua capacidade de generalização e otimização do tempo (Oliveira et al., 2006). Estimativas de H_G , por meio dos modelos de AM, têm sido estudadas e comparadas com modelos estatísticos e os resultados têm 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). Na região Sudeste do Brasil existem poucos estudos com sobre modelagem de H_G usando AM, portanto, o ajuste e a validação dos modelos de AM na estimativa de H_G é de grande interesse para o mapeamento e aproveitamento potencial da radiação solar em projetos em território brasileiro. No presente trabalho é apresentado um estudo comparativo entre o modelo estatístico de Angstrom-Prescott (A-P)^m e modelos AM^m.

METODOLOGIA

Instrumentação e medidas

Os dados utilizados neste trabalho foram medidos na Estação de Radiometria Solar na Faculdade de Ciências Agrárias (FCA) da Universidade Estadual Paulista (UNESP), localizada na cidade de Botucatu (22°53'S latitude, 48°26'W longitude e 786m de altitude). Botucatu é um município brasileiro localizado na região centro-oeste do estado de São Paulo, com uma área total de 1482.642 km². A cidade tem elevado gradiente de altitude entre 400 e 500m na região mais baixa e entre 700 e 900 na região serrana. Esta diferença provoca variações na temperatura do ar e nos ventos. Com bioma cerrado e de mata atlântica, segundo a classificação climática de Köppen o clima da região de Botucatu é Cwa, caracterizado pelo clima tropical de altitude, verão quente e úmido com elevada precipitação, inverno seco e temperatura média do mês mais quente superior do que 22°C (Santos e Escobedo, 2016).

Foram utilizados dados de duração do brilho solar, precipitação, temperatura (máxima e a mínima) e umidade relativa do ar do período de 1996 a 2011. A irradiância solar global (I_G , W m⁻²), foi monitorada por um piranômetro Eppley PSP com erro 4,1% (Reda et al., 2008). Na aquisição dos dados de I_G foi utilizado um datalogger CR23X 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 da irradiação integrada no dia (H_G) (Chaves e Escobedo, 2000). Os dados de brilho solar (n, horas), foram obtidos por um heliógrafo Campbell-Stokes; a precipitação (P, mm) medida através de um pluviógrafo da marca Ota Keiki Seisakusho; temperatura máxima e mínima do ar (T, °C) medida por termômetro de bulbo de mercúrio e álcool, respectivamente; e a umidade relativa do ar (UR, %) medida através de um higrômetro seguindo as recomendações sugeridas pela World Meteorological Organization-WMO (1981).

Modelo de Angstrom-Prescott (A-P)

Vários modelos são sugeridos para estimar H_G utilizando a duração do brilho solar como variável de entrada. O modelo mais conhecido de estimativa da H_G foi proposto por Angstrom (1924) e posteriormente modificado por Prescott (1940), por meio da Equação (1): onde (H_o) é a irradiação solar no topo da atmosfera, (n) é o brilho solar e (N) é o fotoperíodo:

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

O coeficiente “a” pode ser interpretado como a fração de H_G que atinge a superfície da Terra num dia nublado, sendo dependente do tipo e espessura de nuvens. O coeficiente “b” é um complemento que dá o total de H_G . A soma (a+b), é a fração potencial de irradiação solar no topo da atmosfera disponível para alcançar a superfície (isto é, H_G num dia de céu claro).

Support Vector Machine (SVM) com o algoritmo Sequential Minimal Optimization (SMO)

Máquina de Vetor de Suporte (SVM) é uma técnica de aprendizado supervisionado, baseado na teoria de aprendizado estatístico (Vapnik, 1995). Informações mais detalhadas sobre SVM podem ser encontradas em Vapnik (1998). Devido a sua capacidade de fornecer excelente desempenho de generalização, a SVM tornou-se uma poderosa ferramenta para resolver problemas de reconhecimento de padrões, classificação, previsão e de regressão (Shevade et al., 2000). A solução de problemas de regressão usando a SVM pode ser

dada através de um algoritmo iterativo, chamado de Sequential Minimal Optimization (SMO) (Smola e Schölkopf, 1998).

Multilayer Perceptron (MLP) com algoritmo Back Propagation (BP)

A ANN utilizada foi a Multilayer Perceptron (MLP). A MLP mapeia conjuntos de dados de entrada para um conjunto de dados de saída. Essa técnica é amplamente utilizada em modelagem para resolução de problemas complexos.

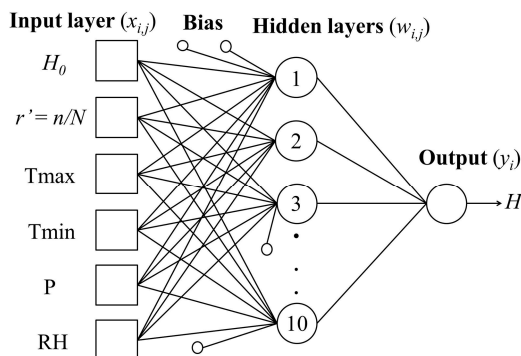


Figura 1. Diagrama em blocos demonstrando a arquitetura da Artificial Neural Network (ANN) utilizada, adaptado de Haykin (1998).

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\$), (Lyra et al., 2015). Os sinais de entrada são enviados para a camada oculta. Em seguida, as camadas ocultas e de saída multiplicam os sinais de entrada por um conjunto de pesos.

A MLP típica, com uma camada oculta, pode ser modelada conforme Equação (2) (Lam et al., 2008b):

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

Onde \$\theta_i\$ é a bias do neurônio \$i\$. Cada entrada é multiplicada por um peso de ligação. A saída dos neurônios é calculada pela aplicação de uma função de ativação não linear, Equação (3), que é tipicamente sigmóide padrão (Rehman e Mohandes, 2008).

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

Existem muitos algoritmos de aprendizado específicos para determinados modelos de redes neurais. A MLP foi treinada usando o algoritmo de aprendizagem supervisionado Back Propagation (BP) e o termo momento. Neste algoritmo o valor de cada camada de saída é usado para atualizar o peso da camada anterior. O BP aprende de forma iterativa do processamento conjunto de exemplos de treinamentos de dados. O ajuste de peso na iteração depende da taxa de aprendizado e do momentum. A taxa de aprendizagem durante cada interação controla o tamanho das mudanças de peso e do viés.

Software utilizado

A caixa de ferramentas do Waikato Environment for Knowledge Analysis (WEKA) foi usada para treinar e validar os dados de \$H_G\$ com o algoritmo SMO para SVM e a BP para ANN. WEKA consiste num conjunto de algoritmos de AM, contendo ferramentas para pré-processamento de dados, classificação, regressão, regras de associação e visualização (Witten et al., 2011). O SMO é usado com a função de Kernel RBF para formação dos modelos. Nos modelos de ANN, com o algoritmo BP, foram considerados os valores: taxa de aprendizado = 0,3; momentum = 0,2 e número de interações = 500. As camadas ocultas (*hidden layers*) 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 “ α ” = [(variáveis de entrada + classes)/2]. A função de ativação sigmoidal é adotada.

Base de dados de validação dos modelos: seleção dos anos típico e atípico

Na validação dos modelos gerados neste trabalho foi utilizada uma base dados de 2 anos da irradiação solar global (H_G), denominadas de ano típico e atípico, separados a partir da base de dados total de 16 anos (1996 a 2011). A seleção dos anos típico e atípico, foi efetuada através de análises estatísticas dos dados onde se comparou para cada mês do ano, o valor da irradiação solar global (H_G) média inter-anos com desvio padrão de cada mês e o valor da irradiação média de cada mês por ano. O critério de seleção do ano típico é semelhante ao processo de seleção do ano meteorológico típico divulgado pela *World Meteorological Organization* (WMO) em 1981. Os resultados encontrados para os demais meses estão representados na Tabela 1, a qual mostra a constituição mensal dos anos típico e atípico do banco de dados total de 16 anos.

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

		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

Indicativos Estatísticos

Existem vários índices estatísticos utilizados para avaliar o desempenho de modelos. Na performance dos modelos de radiação solar com técnicas de aprendizados de maquina normalmente são: Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE), coeficiente de correlação (r), Coeficiente de determinação (R^2) (Santos et al., 2016). Uma escala classificatória para os diferentes intervalos de rRMSE para avaliar o desempenho dos modelos é utilizada (Jamieson et al., 1991; Li et al., 2013): Excelente se $rRMSE < 10\%$; Bom se $10\% \leq rRMSE < 20\%$; Aceitável se $20\% \leq rRMSE < 30\%$; Pobre se $rRMSE \geq 30\%$.

RESULTADOS E DISCUSSÃO

Correlação entre a transmissividade atmosférica da irradiação solar global (H_G/H_0) e a razão de insolação (n/N): Modelo de Angstrom-Prescott (A-P)

A Figura. 2 mostra a correlação entre a transmissividade atmosférica da irradiação solar global (H_G/H_0) e a razão de insolação (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 semelhante à maioria dos trabalhos da equação de Angstrom-Prescott em diversos países (Martinez-Lozano, 1984; Akinoglu, 1990; Bakirci, 2009a).

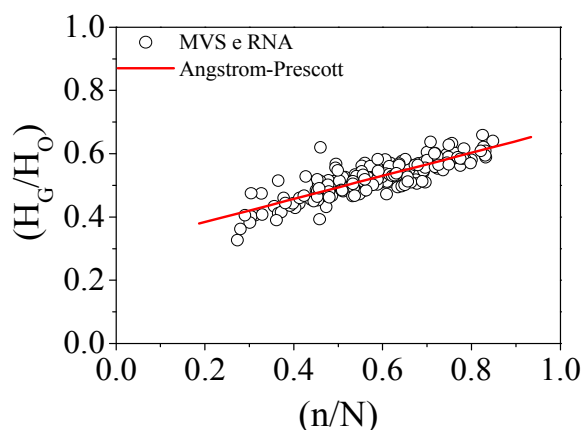


Figura 2. Correlação entre a transmissividade atmosférica da irradiação solar global (H_G/H_0) e a razão de insolação (n/N), e a reta obtida por regressão linear no período de 1996-2011.

A Equação (4) de (A-P) obtida na correlação da Figura 2, por regressão linear, foi:

$$(H_G/H_0) = 0,311 + 0,366 \times (n/N) \quad (4)$$

O coeficiente de determinação (R^2) = 0,714, bem como o valor da fração de H_G mínima ($a = 0,311$), do complemento que dá o valor total de H_G ($b = 0,366$) ou a fração atmosférica máxima ($a+b = 0,677$) diferem dos valores obtidos no intervalo de variação dos coeficientes em diversas localidades do Brasil ($a = 0,248 \pm 0,081$), ($a+b = 0,689 \pm 0,124$) e R^2 cuja média é de 0,793 ($\pm 0,162$) (Tabela 2). Fatores climáticos, como

tipo e espessuras de nuvens, concentração de vapor de água e aerossóis na atmosfera são os fatores responsáveis pelas variações de a e $a+b$ em cada local.

Tabela 2. Coeficientes de Angstrom-Prescott (A-P) obtidos no território brasileiro por diversos autores.

Autores	Localidades (Nº)	Coeficientes de (A-P)			
		a	b	$a+b$	R^2
Tiba (2001)	34 (NE)	0,22 - 0,35	0,31 - 0,58	0,58 - 0,84	0,71 - 0,98
Ribeiro et al. (1982)	Manaus (AM)	0,26	0,48	0,74	0,69
Souza et al. (2016)	3 (AL)	0,16 - 0,39	0,28 - 0,58	0,68 - 0,75	*
Tarifa (1972)	Presidente Prudente (SP)	0,19	0,39	0,58	*
Cervellini et al. (1966)	2 (SP)	0,19 - 0,23	0,56 - 0,61	0,79 - 0,80	*
Modelo proposto (2015)	Botucatu (SP)	0,31	0,37	0,68	0,71

Siglas: AL = Alagoas, AM = Amazonas, SP = São Paulo, NE = Região Nordeste do Brasil.

- Informação não fornecida pelos autores.

A Figura. 3 (a, b) mostra as correlações obtidas na validação entre as estimativas e as medidas de H_G , para as bases de dados típico e atípico. As distribuições dos valores estimados pela equação (A-P) e as medidas nas duas condições de validação estão em concordância linear com a reta ideal (45°) ou $r = 1$ no ajuste. As retas de regressão linear com os coeficientes de correlação iguais $r^t = r^a = 0,980$ mostram que a equação (A-P) pode estimar H_G com elevados coeficientes de determinação, próximos de 1: $(R^2)^t = 0,960$ (Fig. 3 a) e com $(R^2)^a = 0,960$ (Fig. 3 b).

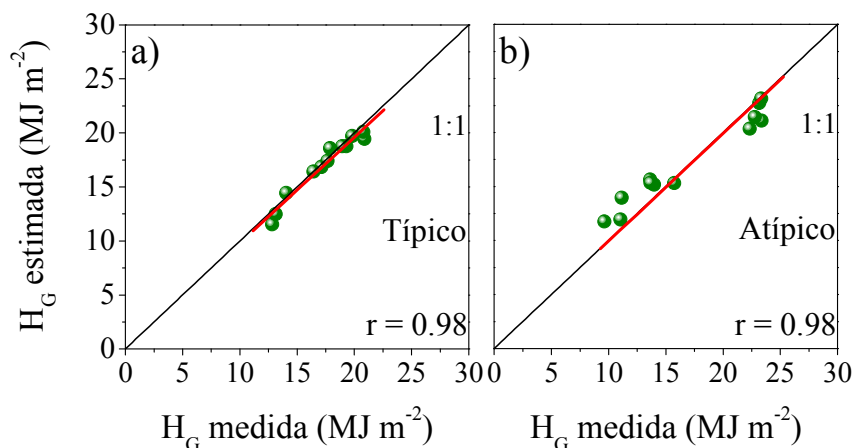


Figura 3 (a, b). Comparação entre os valores estimados por meio de Angstrom-Prescott (A-P) e medidas de H_G para anos típico e atípico.

O indicativo estatístico RMSE obtidos na validação entre as estimativas e as medidas de H_G , para as bases de dados típico e atípico foram: $RMSE^t = 3,9\%$ e $RMSE^a = 9,8\%$.

Treinamento e validação dos modelos das técnicas de SVM e ANN

O mesmo banco de dados da geração do modelo de A-P (Figura 2) foi usado no treinamento das técnicas SVM e ANN. O Modelo de Angstrom-Prescott e os modelos de SVM1 e ANN1 (combinação1) utilizam as mesmas variáveis de entrada H_0 , N (calculados) e n (medido). Além da combinação1, foram treinadas três novas combinações (SVM2, SVM3 e SVM4; ANN2, ANN3 e ANN4), conservando as mesmas variáveis da combinação 1, e acrescentando na sequência as seguintes variáveis de entrada: temperatura máxima e mínima do ar (T , °C), precipitação (P , mm), Umidade Relativa (UR , %), como mostra a Tabela 3.

Tabela 3. Variáveis de entrada para os modelos de SVM e ANN.

Modelo	SVM	ANN	Variáveis de entrada
Combinação 1	SVM1	ANN1	H_0 , n/N
Combinação 2	SVM2	ANN2	H_0 , $T_{máx}$, $T_{mín}$, n/N
Combinação 3	SVM3	ANN3	H_0 , $T_{máx}$, $T_{mín}$, P, n/N
Combinação 4	SVM4	ANN4	H_0 , $T_{máx}$, $T_{mín}$, P, UR, n/N

H_0 = irradiação solar no topo da atmosfera, n/N = razão de insolação, $T_{máx}$ = temperatura máxima do ar, $T_{mín}$ = temperatura mínima do ar, P = precipitação pluvial e UR = umidade relativa do ar.

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

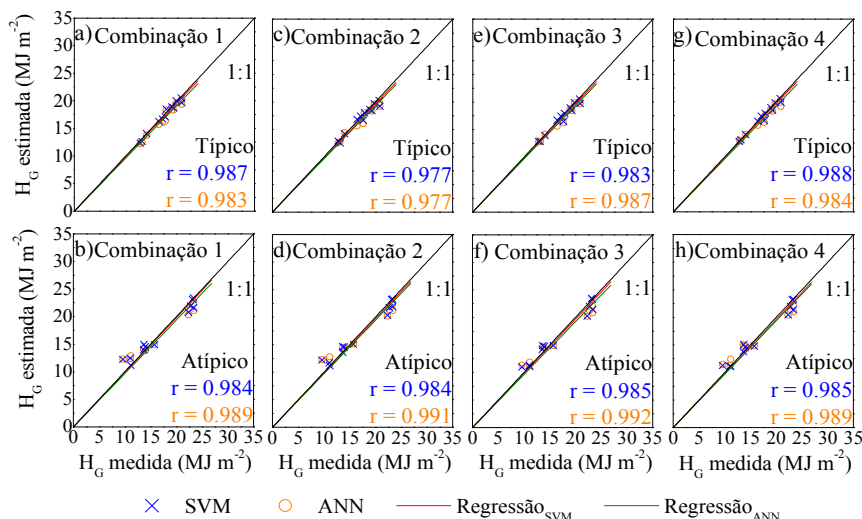


Figura 4 (a-h). Correlação entre os valores estimados pelos modelos de AM, e os valores medidos de H_G .

Os valores estimados por SVM e ANN nas quatro (4) combinações estão em concordância linear com as medidas Figuras 4 (a, b): os valores de coeficientes de correlação (r) obtidas nas regressões lineares estão muito próximos da reta ideal (45°) ou $r = 1$ no ajuste.

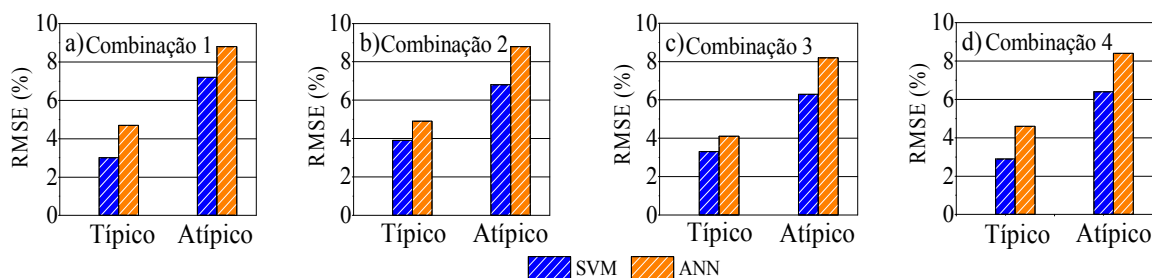


Figura 5 (a-d). Indicativo estatístico RMSE (%) para anos típico e atípico.

Para a combinação 1 (SVM1 e ANN1), os valores de $r^t = 0,987$ e $r^a = 0,984$; $r^t = 0,983$ e $r^a = 0,989$, respectivamente, mostram que a medida e a estimativa de H_G estão bem correlacionadas para todas as combinações. A técnica SVM1 apresentou maior valor de r que a rede ANN1, na validação do ano típico e menor no atípico. O indicativo estatístico rRMSE, obtido 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 Fig. 4: $rRMSE^t_{SVM1} = 3,0\%$ e $rRMSE^a_{SVM1} = 7,2\%$, enquanto para $rRMSE^t_{ANN1} = 4,7\%$ e $rRMSE^a_{ANN1} = 8,8\%$.

Comparação do desempenho entre os modelos Angstrom-Prescott (A-P) e as redes MVS e RNA mensal em Botucatu e de outros locais

A Tabela 4 mostra a comparação dos indicativos estatísticos RMSE, RrMSE, em unidades de energia por mês e porcentagem, e, também, os valores de r e R^2 , obtidos em duas condições validações, típico e atípico, entre medida e estimativa dos modelos estatístico de Angstrom-PreScott e das AM's MVS e RNA (mensal) em Botucatu, bem como para outros locais no globo terrestre. Os valores dos indicativos estatísticos rRMSE, RMSE e R^2 do modelo de A-P e das técnicas de AM (MVS e RNA), obtidas em Botucatu, mostram que os modelo MVS1^m apresenta melhor resultado que o modelo estatístico (A-P)^m e o RNA1^m nas duas condições de validação, respectivamente.

Tabela 4. Comparações entre os modelos estatísticos de A-P com as técnicas de aprendizado de máquina (AM).

Autores	Localidades (N°)	Modelos	rRMSE	RMSE	R^2
			(%)	(MJ m ⁻²)	
Mohammadi et al. (2015a) ^m	Isfahan (Irã)	A-P	6,0	1,16	0,96
Mohammadi et al. (2015a) ^m	Isfahan (Irã)	MVS	2,26	0,45	0,99
Tymvios et al. (2005) ^m	Nicósia (Chipre)	A-P	13,36	-	0,85
Tymvios et al. (2005) ^m	Nicósia (Chipre)	RNA	5,67 a 10,15	-	0,74 a 0,91
Modelo mensal proposto	Botucatu (Brasil)	A-P^m	3,94 a 9,76	0,67 a 1,66	0,96 a 0,96
Modelo mensal proposto	Botucatu (Brasil)	MVS1^m	3,0 a 7,2	0,51 a 1,22	0,96 a 0,98
Modelo mensal proposto	Botucatu (Brasil)	RNA1^m	4,7 a 8,8	0,82 a 1,50	0,96 a 0,98

A comparação entre SVM1 x (A-P) mostra que os valores de $rRMSE_{SVM1}^t = 3,0\%$ e $rRMSE_{SVM1}^a = 7,2\%$ são inferiores aos de $rRMSE_{(A-P)}^t = 3,94\%$ e $rRMSE_{(A-P)}^a = 9,76\%$ (Fig. 6 b); e os de $(R^2)_{SVM1}^t = 0,974$ e $(R^2)_{SVM1}^a = 0,968$ são superiores aos de $(R^2)_{(A-P)}^t = 0,953$ e $(R^2)_{(A-P)}^a = 0,960$. Os resultados obtidos para o indicativo estatístico rRMSE para SVM1 e (A-P): neste trabalho estão de acordo com os resultados dos trabalhos por Mohammadi et al. (2015a) em Isfahan no Irã com $rRMSE_{SVM1} = 2,3\%$ e $rRMSE_{(A-P)} = 6,0\%$ e valores de $R^2_{SVM1} = 99,0\%$, e $R^2_{(A-P)} = 96\%$.

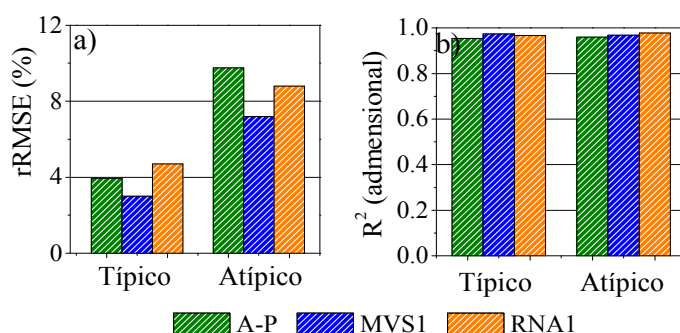


Figura 6 (a, b). Indicativos estatísticos rRMSE e R^2 para anos típico e atípico.

A comparação entre (A-P) x ANN1 mostra que: os valores de $rRMSE_{(A-P)} = 3,94\% - 9,76\%$ são inferiores a $rRMSE_{ANN4} = 4,7\% - 8,8\%$, os valores de $R^2_{(A-P)} = 96,0\% - 96,0\%$ são inferiores a $R^2_{ANN4} = 96,0\% - 98,0\%$. Os resultados obtidos neste trabalho para os indicativos estatísticos $rRMSE_{(A-P)} = 3,94\%$ a $9,76\%$; $R^2_{(A-P)}$ de 96% a 96% ; $rMBE_{ANN4} = -3,8\% - 0,9\%$; $rRMSE_{ANN4} = 4,7\%$ a $8,8\%$; e R^2_{ANN4} de 96% a 98% são na mesma ordem de grandeza aos de Tymvios et al. (2005) em Nicósia (Chipre), cujos resultados para os indicativos estatísticos $rRMSE_{(A-P)} = 13,36\%$; $rRMSE_{ANN4} = 5,67\%$ a $10,15\%$.

CONCLUSÃO

A equação de estimativa (A-P), obtidas por meio de regressão linear em Botucatu, SP, com coeficientes de determinação $R^2 = 0,715$ é da mesma ordem de grandeza aos valores de R^2 obtidos por outros pesquisadores e outros locais do globo terrestre. Os coeficientes de correlação, $r = 0,976$ e $r = 0,980$, para validação com o ano típico e atípico, mostram que o modelo de A-P pode estimar H_G com elevados coeficientes de determinação.

O treinamento das técnicas SVM1 e a ANN1 com mesmas variáveis de entrada H_0 e n/N que o modelo A-P com valores de r próximos dos 100% mostram que as redes podem estimar H_G na mesma ordem de grandeza que a equação de Angstrom-Preseott clássica.

Na comparação entre desempenho para os modelos A-P, SVM1 e ANN1 por meio dos valores dos indicativos estatísticos $rRMSE$, $RMSE$, r e R^2 obtidos da validação mostrou que: entre o modelo A-P e a técnica SVM1 em Botucatu, a técnica SVM1 apresentou melhor resultado que o modelo estatístico de A-P; Entre os modelos SVM1 e ANN1 em Botucatu a técnica SVM1 apresentou melhor resultado do que a rede ANN1; entre o modelo A-P e a rede ANN1 em Botucatu o modelo estatístico A-P, apresentou no geral melhor resultado do que a rede ANN1.

REFERÊNCIAS

- Aknoglu, B. G. (1990). A review of sunshine-based models used to estimate monthly average global solar radiation. *Renewable Energy*, 1, 479-497.
- Angstrom, A. (1924). Solar and terrestrial radiation. *Quarterly Journal of the Royal Meteorological Society*, 50, 121-125.
- Bakirci, K. (2009a). Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. *Energy*, 34, 485-501.
- Bakirci, K. (2009b). Models of radiation with hours of bright sunshine: A review. *Renewable and Sustainable*, 13, 2580-2588.
- Bosch, J. L.; López, G.; Batlles, F. J. (2008). Daily solar irradiation estimation over a mountainous area using Artificial Neural Networks. *Renewable Energy*, v. 33, 1622-1628.
- Cervellini, A.; Salati, E. (1966). Estimativa da distribuição da energia solar no Estado de São Paulo. *Bragantia*.
- Chaves, M. A.; Escobedo, J. F. (2000). A software to process daily solar radiation data. *Renewable Energy*, 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*, 75, 311-318.
- Elizondo, D.; Hoogenboom, G.; Mclendon, R. W. (1994). Development of a neural network model to predict daily solar radiation. *Agricultural and Forest Meteorology*, 71, 115-132.
- Escobedo, J. F.; Teramoto, E. T.; Oliveira, A. P. et al. (2012). 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 y Medio Ambiente*, 16.
- Haykin, S. (1998). *Neural networks: A comprehensive foundation*. 2nd ed. pp. 897. Hamilton: Prentice Hall.
- Hsiao, T.; Heng, L.; Steduto, P. et al. (2008). Aqua Crop – The FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. *Agronomy Journal*, 101, 448-459.
- 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*, 27, 337-350.
- Li, M.; Tang, X.; Wu, W. et al. (2013). General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Conversion and Management*, 70, 139-148.
- Lam, J. C.; Wan, K. K. W.; Yang, L. (2008). Solar radiation modeling using ANNs for different climates in China. *Energy Conversion & Management*, 49, 1080-1090.
- Martinez-Lozano, J. A.; Tena, F.; Onrubia, J. E. et al. (1984). The historical evolution of the Angstrom Formula and its modifications. *Agricultural and Forest Meteorology*, 33, 109-128.
- Nicácio, R. M.; Souza, J. L.; Bernardo, S. O. (2001). Estimation da irradiância solar global para Maceió utilizando o modelo linear de Angstrom-Preseott. In: XII Congresso Brasileiro de Agrometeorologia – Fortaleza.
- Oliveira, A. P.; Escobedo, J. F.; Machado, A. J. et al. (2002). Correlation model of diffuse solar-radiation applied to the city of São Paulo, Brazil. *Applied Energy*, 71, 59-73.
- Oliveira, A. P.; Soares, J.; Boznar, M. Z.; Mlakar, P.; Escobedo, J. F. (2006). An application of Neural Network technique to correct the dome temperature effects on pyrgeometer measurements. *Journal of Atmospheric and Oceanic Technology*, 23, 80-89.
- Mohammadi, K.; Shamshirband, S.; Anisi, M. H. et al. (2015). Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion and Management*, 91, 433-441.
- Piri, J.; Shamshirband, S.; Petkovic, D. et al. (2015). Prediction of the solar radiation on the Earth using support vector regression technique. *Infrared Physics & Technology*, 68, 179-185.

- Prescott, J. A. (1940). Evaporation from a water surface in relation to solar radiation. Transactions of the Royal Society Science Australian, 64, 114-118.
- 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, 3, 58-66.
- Rehman, S.; Mohandes, M. (2008). Artificial neural network estimation of global solar radiation using air temperature and relative humidity. Energy Policy, 36, 571-576.
- Ribeiro, M. N. G.; Salati, E.; Vila Nova, N. A.; Demétrio, C. G. B. (1982). Radiação solar disponível em Manaus (AM) e sua relação com a duração do brilho solar. Acta Amazônica, 12, 339-346.
- Shevade, S. K.; Keerthi, S. S.; Bhattacharyya, C. (2000). Improvements to the SMO Algorithm for SVM Regression. IEEE Transactions on Neural Networks, v. 11.
- Smola, A. J.; Schölkopf, B. (1998). "A Tutorial on Support Vector Regression". Royal Holloway College, London, U. K., Neuro COLT Tech. Rep. TR 1998-030.
- 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, Applied Energy, 79, 201-214.
- Santos, C. M.; Escobedo, J. F. (2016). Temporal variability of atmospheric turbidity and DNI attenuation in the sugarcane region, Botucatu/SP. Atmospheric Research, 181, 312-321.
- Santos, C. M.; Escobedo, J. F.; Tadao, E. T.; Silva, S. H. M. G. (2016). Assessment of ANN and SVM models for estimating normal direct irradiation (H_b). Energy Conversion and Management, 126, 826-836.
- Souza, J. L.; Lyra, G. B.; Santos, C. M. et al. (2016). Empirical models of daily and monthly global solar irradiation using sunshine duration for Alagoas State, Northeastern Brazil. Sustainable Energy Technologies and Assessment, 14, 35-35.
- Tarifa, J. R. (1972). Estimativa da radiação em função da insolação para Presidente Prudente. Caderno de Ciências da Terra, 22, 1-15.
- Tiba, C. (2001). Solar radiation in the Brazilian Northeast. Renewable Energy, 22, 565-578.
- Tymvios, F. S.; Jacovides, C. P.; Michaelides, S. C. et al. (2005). Comparative study of Angstrom and artificial neural networks methodologies in estimating global solar radiation. Solar Energy, 78, 752-762.
- Vapnik, V. N. (1995). The nature of Statistical learning theory. Springer-Verlag, New York.
- Vapnik V, N. (1998). Statistical learning theory. New York: Wiley.
- World Meteorological Organization. Meteorological Aspects of the utilization of Solar Radiation as an Energy Source. (1981). World Meteorological Organization Technical Note No. 172, WMO-No. 557, Geneva, pp. 298.
- Witten, I. H.; Frank, E.; Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques. 3 rd ed. pp. 630.

Estimation of Monthly Global Solar Irradiation by Models of Angstrom -Prescott and Machine Learning in Botucatu/SP/Brazil

ABSTRACT: In this paper we describe the comparative study between the Angstrom-Prescott (AP) model with two machine learning techniques [Support Vector Machine (SVM) and Artificial Neural Network (ANN)], in the estimation of monthly global solar irradiation (H_G). The H_G database used was measured in the period from 1996 to 2011 in Botucatu/SP/Brazil. The (A-P) model obtained $(H_G/H_0) = 0,311 + 0,366 * (n/N)$, with $R^2=0.710$ was validated using two databases: typical (t) and atypical (a). The statistical indicatives r , $rRMSE$, obtained in the comparison between the estimation and measurement were: $r_t=r_a=0.980$, $rRMSE_t=3.9\%$ and $rRMSE_a=9.8\%$. The SVM and ANN techniques were trained in four combinations of inputs of meteorological variables and validated in the databases: typical and atypical. The comparison shows that: the SVM has better performance than the (A-P) model and the ANN in estimating H_G .

Keywords: Solar radiation, Angstrom-Prescott, artificial intelligence.