

ESTIMATIVA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL USANDO REDES NEURAS ARTIFICIAIS E COMPARAÇÃO COM MODELOS ESTATÍSTICOS

C. M. dos Santos¹, J. F. Escobedo², S. H. M. G. da Silva³, T. J. Rossi⁴, M. B. P. da Silva⁵
^{1,2,4,5}Grupo de Radiometria Solar, Departamento de Engenharia Rural –
FCA/UNESP/Botucatu/SP/Brasil. Tel. +55143880-7585. e-mail: ciceromanoel@fca.unesp.br
³UNESP Campus Experimental de Registro/SP/Brasil.

Recibido 18/09/15, aceptado 14/10/15

RESUMO: A irradiação direta na incidência normal (H_b) foi estimada utilizando Redes Neurais Artificiais (ANN2) para uma base de dados medida do período de 1996 a 2008 em Botucatu- Brasil, região com influência de queimadas agrícolas. A rede utilizada foi a Multilayer Perceptron (MLP), com o algoritmo de aprendizado *backpropagation* e *termo momentum*. A ANN2 foi comparada com um modelo estatístico de ajuste local (ME4) e dois modelos selecionados na literatura (ME5 e ME6). A variável de entrada foi o k_t (transmissividade atmosférica da irradiação global). Os modelos foram avaliados por meio dos indicativos estatísticos: Mean Bias Error (MBE, MJm^{-2}), Relative Mean Bias Error (rMBE, %), Root Mean Square Error (RMSE, MJm^{-2}), Relative Root Mean Square Error (rRMSE, %) e d de Willmott. O desempenho de ANN2 é melhor que dos modelos estatísticos. A estimativa com ANN2 resultou: rRMSE=16,519% e $d=0,980$. O ME4 estimou com rRMSE=20,130% e $d=0,971$. Os ME5 e ME6 tiveram desempenho inferior ao ME4. O modelo ANN2 é recomendado como a primeira escolha para estimativa de H_b , seguido do modelo ME4.

Palavras-Chave: Radiação Solar, WEKA, Mineração de Dados, Aprendizado de Máquinas, Modelagem.

INTRODUÇÃO

Com a demanda energética mundial e a necessidade de novas fontes alternativas de energia renovável, a irradiação direta na incidência normal (H_b) é destaque em estudos para sistemas de conversões de energia solar. Os sistemas de conversões solares consistem em elétrico e térmico. A conversão da energia solar em elétrica é feita através de células fotovoltaicas e sistemas de concentradores solares. Nos sistemas concentrados a H_b é destaque por ser a única componente com potencial para ser concentrada com elevado grau de aproveitamento. Na conversão solar em térmica a H_b é aplicada no aquecimento e iluminação de ambientes, nos processos de geração de calor e no aquecimento de água.

Portanto, valores de H_b são importantes em muitas áreas de pesquisas. No entanto, dados de H_b em superfície não estão facilmente disponíveis na maioria das estações solarimétricas. Esta escassez, principalmente nos países pobres e subdesenvolvidos, é causada pelo elevado preço de importação e manutenção periódica dos sensores (Santos et al., 2014a). O Brasil, inserido entre os países emergentes, não dispõe de muitos sensores destinados ao monitoramento de H_b . Escassos são os centros de pesquisas que monitoram periodicamente. Com a demanda mundial de dados solarimétricos, modelos têm sido sugeridos e utilizados com bastante frequência para estimativa de H_b em todo mundo. Estimativas precisas de H_b dependem da variabilidade temporal e espacial.

¹ Físico, Doutorando em Irrigação e Drenagem FCA/UNESP/BOTUCATU.

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

³ Prof. Dr. UNESP-Registro

⁴ Graduando UNESP-Botucatu.

⁵ Eng. Agrônomo, Mestrando em Irrigação e Drenagem FCA/UNESP/BOTUCATU.

Assim como, dependem das condições climáticas, que são determinantes para variação de H_b e instalação de sistemas de energia solar. Por isso, é recomendado selecionar adequadamente o modelo para aplicação no local de interesse. Muitos são as modelagens utilizadas para geração de base de dados de H_b com boa aceitação.

Na literatura existem muitos modelos estatísticos para estimativa de H_b (Vignola e McDaniels, 1986; Louche et al., 1991; Nfaoui e Buret, 1993; López et al., 2000). Estes modelos são baseados nas metodologias de (Liu e Jordan, 1960) e (Ångström, 1924) e têm sido aplicados em muitos locais. Para lidar com as limitações dos modelos estatísticos, técnicas de inteligência artificial, como as Redes Neurais Artificiais (ANNs) estão sendo aplicadas com sucesso para projetos de previsão de irradiação solar e sistemas solares (Qazi et al., 2015). Koca et al., (2011) aplicaram ANNs na estimativa da H_g em sete locais da região do mediterrâneo de Anatólia. De acordo com os resultados, o modelo é potencialmente indicado para estimar H_g em qualquer região com erros pequenos e aceitáveis. Em oito estações no Omã, Reddy (2003), definiu que a rede Função de Base Radial pode ser utilizada para estimar a H_g em qualquer local. Para estimar a irradiação solar global média mensal na Turquia, (Ozgoren et al., 2012), desenvolveram um modelo de ANNs baseado na regressão não linear múltipla. Os resultados obtidos para os erros estão dentro de limites aceitáveis e a ANNs pode ser usada para estimativa. Wana et al., (2008) utilizaram regressão e ANNs para investigar as correlações entre a radiação solar global diária e a insolação para locais de climas diferentes na China.

Em estudos para a cidade de São Paulo, Soares et al., (2004), utilizaram ANNs na estimativa da irradiação solar difusa e compararam os resultados com modelos estatísticos. De acordo com os autores a ANNs teve melhor desempenho e pode ser utilizada na modelagem solar com resultados satisfatórios. Jiang (2009) utilizou ANNs para estimar irradiação solar global em oito locais da China. As estimativas tiveram boas concordâncias e possuem desempenhos estatísticos melhores que os modelos empíricos. Kalogirou et al., (2014) empregaram ANNs na previsão de grandes sistemas solares. Karatepe et al., (2006) compararam ANNs com modelos estatísticos para modelagem de sistemas fotovoltaicos. Os resultados de comparação foram satisfatórios e a ANNs pode ser usado com sucesso. Zarzalejo et al., (2005) utilizaram técnicas de inteligência artificial, lógica Fuzzy e ANNs, na estimativa da radiação global horário por imagens de satélite em 15 locais da Espanha. Em oito locais da região Nordeste do Brasil, Lyra et al., (2015), utilizaram ANNs e modelos empíricos para estimativa da irradiação solar global. Apesar dos inúmeros trabalhos com ANNs na estimativa de H_g (Yadav e Chandel, 2014), escassos são os trabalhos na estimativa de H_b utilizando ANNs.

Para facilitar o uso por usuários, as técnicas de aprendizagem de máquinas, tal com ANNs estão sendo disponibilizados em Softwares. O Waikato Environment for Knowledge Analysis (WEKA) é um conjunto de algoritmos de AMs, inclusive com a ANNs, amplamente utilizado com sucesso em diversas áreas (Oliveira et al., 2006; Lorena et al., 2011).

O objetivo deste trabalho é estimar a H_b utilizando Redes Neurais Artificiais. A rede neural é comparada com modelos estatísticos na partição de tempo diário. A rede possui a mesma variável de entrada dos modelos estatísticos, a transmissividade atmosférica da irradiação global (k_t). A ANNs utilizada foi a Multilayer Perceptron (MLP) implementada no WEKA. O estudo está organizado da seguinte forma. A descrição dos dados, local de estudo e metodologia adotada é dada na primeira parte. Nos resultados é mostrado o desempenho dos modelos, destacando a metodologia com melhor desempenho. Por último as principais conclusões são destacadas.

MATERIAL E MÉTODOS

Local de estudo e dados.

Os dados utilizados foram medidos na Estação Radiométrica localizada na Faculdade de Ciências Agronômicas da UNESP/Botucatu (latitude 22,85°S, longitude 48,45°W e altitude 786m) (Figura 1a). 13 anos (1996 a 2008) de dados irradiação direta na incidência normal, irradiação solar global (H_g) e insolação solar separados em partições de tempo foram utilizadas. A H_b foi obtida por pireliômetro da Eppley NIP acoplado a um rastreador solar ST3 da Eppley, a H_g foi monitorada por piranômetro Eppley PSP e a insolação (n) solar foi obtida por heliógrafo convencionais do tipo Campbell-Stokes

(Figura 1b, c, d). Os sensores na estação radiométrica são aferidos anualmente pelo método comparativo sugerido pela Organização Mundial de Meteorologia (WMO, 2008).



Figura 1: Local e dados utilizados. a) localização da estação, b) Pireliômetro, c) Piranômetro e d) Heliógrafo.

A região de Botucatu é caracterizada por ser de clima temperado quente (mesotérmico), verão quente e úmido com elevada precipitação e inverno seco (Escobedo et al., 2011). A Tabela 1 apresenta as séries climáticas da temperatura do ar, umidade relativa do ar e precipitação acumulada, do período de 1971 a 2009. O mês mais quente é fevereiro ($22,53 \pm 0,72^\circ\text{C}$), o mais frio julho ($16,77 \pm 0,89^\circ\text{C}$) e a temperatura do ar média anual de $20,086 \pm 2,176^\circ\text{C}$. A umidade relativa do ar varia de $61,15 \pm 5,92\%$ (agosto) a $76,41 \pm 2,54\%$ (janeiro), com média anual de $69,00 \pm 4,633\%$. A precipitação média anual acumulada é de $1381,29$ mm, com chuvas mensais entre $38,88 \pm 37,80$ mm (junho) e $305,75 \pm 120,03$ mm (janeiro). O período mais chuvoso ocorre de outubro a março (primavera-verão) e o menos chuvoso entre abril e setembro. No período chuvoso, a precipitação é causada pela Zona de Convergência do Atlântico Sul (ZCAS). No período seco, a precipitação é do tipo frontal originada do encontro das massas de ar frias e secas vindas da região sul com as massas quentes e úmidas, da região sudeste.

Mês	T_{ar} ($^\circ\text{C}$)	UR (%)	P (mm)
Jan	$22,13 \pm 0,81$	$76,41 \pm 2,54$	$305,75 \pm 120,03$
Fev	$22,53 \pm 0,72$	$75,15 \pm 1,38$	$188,38 \pm 79,54$
Mar	$22,19 \pm 0,55$	$72,81 \pm 2,51$	$143,38 \pm 61,25$
Abr	$20,96 \pm 1,03$	$70,48 \pm 2,23$	$58,82 \pm 34,98$
Mai	$17,22 \pm 0,99$	$67,90 \pm 2,00$	$75,54 \pm 45,95$
Jun	$17,05 \pm 0,65$	$66,73 \pm 2,38$	$38,88 \pm 37,80$
Jul	$16,77 \pm 0,89$	$63,13 \pm 2,94$	$41,33 \pm 48,08$
Ago	$18,80 \pm 1,12$	$61,15 \pm 5,92$	$38,83 \pm 33,45$
Set	$19,28 \pm 0,83$	$64,76 \pm 4,28$	$70,65 \pm 51,65$
Out	$20,71 \pm 1,07$	$68,48 \pm 2,84$	$109,46 \pm 52,11$
Nov	$21,23 \pm 0,81$	$69,64 \pm 2,68$	$113,60 \pm 61,51$
Dez	$22,17 \pm 0,74$	$71,37 \pm 3,33$	$196,67 \pm 44,40$

Tabela 1: Séries climáticas de a) temperatura e umidade (b) precipitação no período de 1971 a 2009. T_{ar} = temperatura do ar; UR = Umidade relativa do ar e P = precipitação.

Modelos Estatísticos.

Os modelos estatísticos utilizados neste trabalho, afim de comparação, possuem relação entre a fração transmitida da irradiação direta na incidência normal (kt_b) [razão entre a irradiação direta na incidência normal (H_b) e a irradiação direta no topo da atmosfera (H_{sc}^d)] com a transmissividade atmosférica (kt) [razão entre a H_g e a irradiação solar global no topo da atmosfera (H_0)] (Liu e Jordan, 1960). Neste trabalho é considerado modelo estatístico ajustado com dados local para comparação com dois modelos estatísticos existentes na literatura (ME4, ME5 e ME6) (Tabela 2).

Relação de kt_b vs kt		
Autor	Símbolo	Modelo
Ajuste local	ME4	$kt_b = -0,123 + 1,768 \times kt - 8,213 \times kt^2 + 16,617 \times kt^3 - 9,584 \times kt^4$
Nfaoui e Buret (1993)	ME5	$kt_b = 0,13 - 0,95 \times kt + 2,02 \times kt^2$ $kt > 0,25$ $kt_b = 0 \quad kt < 0,25$
Vignola e McDaniels (1986)	ME6	$kt_b = 0,02 - 0,280 \times kt + 0,828 \times kt^2 + 0,765 \times kt^3$

Tabela 2: Modelos estatísticos utilizados na comparação.

A Figura 2 mostra o evolução dos modelos ME4, ME5 e ME6 para valores de kt no intervalo de 0 a 1. Observa-se que para $kt > 0,8$ a curva de ME4 decai, enquanto as curvas dos modelos ME5 e ME6 têm comportamento crescente. Isto é devido, primeiramente, na baixa frequência (ou inexistência) de valores de kt acima de 0,77 para região de Botucatu. Neste caso, a curva ajustada segue a tendência dos dados locais, enquanto que a curva de ME5 e ME6 dos locais onde foram ajustados. O segundo aspecto está relacionado ao modelo ajustado. O modelo ME4 ajustado com dados locais teve o maior coeficiente de determinação (R^2) com o polinômio de 4º, enquanto que no modelo ME5 o ajuste foi dividido em duas partições e o modelo ME6 ajustado com polinômio de ordem três, de acordo com o melhor R^2 obtido pelos autores. Por isso, a curva para o ME4 possui comportamento de uma função polinomial de grau quatro. O ME4 também resultou no decaimento da curva para valores de $kt < 0,12$. Neste caso, caberia uma condicional para zerar os valores estimados de kt_b , quando estes fossem negativos. Mas neste estudo, foi julgado não ser necessário.

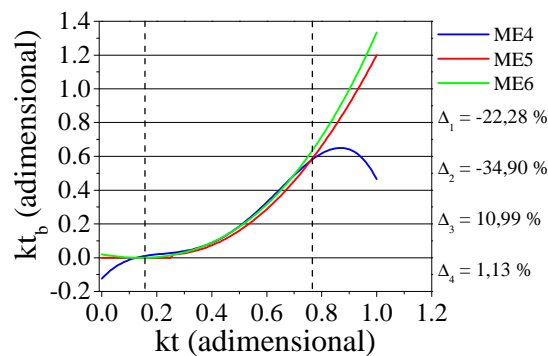


Figura 2: Comportamento dos modelos (ME4, ME5 e ME6) para valores de kt no intervalo de 0 a 1.

Para uma análise mais detalhada calculou-se o desvio (Δ) da curva dos modelos, por meio da relação $\Delta(\%) = [(kt_{bL} - kt_{bF})/kt_{bL}] * 100$, onde kt_{bL} e kt_{bF} são os valores médios da curva para o modelo de ajuste local (ME4) e modelo de ajuste em outro local (ME5 e ME6), respectivamente. Os valores de Δ_1 e Δ_2 são para relação entre (ME4 e ME5) e (ME4 e ME6), respectivamente. Estes desvios foram para o intervalo total de variação de kt (0 a 1) e mostra que as estimativas com ME5 e ME6 são maiores que a estimativa com ME4. Os valores de Δ_3 e Δ_4 são para relação entre (ME4 e ME5) e (ME4 e ME6), respectivamente. Estes desvios foram para o intervalo de kt entre (0,16 e 0,77), destacado pelas duas linhas na vertical. Neste intervalo, o desvio mostra que a o ME4 estima maiores valores de kt_b que os demais modelos e que teoricamente os modelos podem ser utilizados em Botucatu. Nos resultados e discussão, a análise completa do desempenho dos modelos e seu potencial serão apresentados.

Redes Neurais Artificiais (ANNs).

As ANNs foram usadas com sucesso na resolução de problemas complexos em diferentes áreas de aplicação, incluindo o reconhecimento de padrões, identificação, classificação, fala (speech), visão (vision) e sistemas de controle (Hasni et al., 2012).

As ANNs são sistemas computacionais de minerações de dados maciçamente paralelo distribuído, constituído de unidades de processamentos simples que consistem num grupo interconectado de neurônios artificiais que se assemelham na estrutura, no método de processamento e na capacidade de aprendizagem do cérebro humano (Haykin, 1998). Existem diferentes tipos de ANNs, destaca-se o

Perceptron de Múltiplas Camadas (MLP). A MLP, que é amplamente utilizado em modelagem para resolução de problemas complexos, consiste de um conjunto de unidades sensoriais (nós de fonte) com três camadas interligadas. 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) (Figura 3).

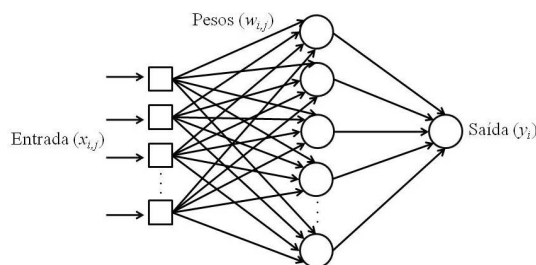


Figura 3: Estrutura de uma MLP com uma camada oculta.

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. A saída dos neurônios é calculada pela aplicação de uma função de ativação. A função de ativação é tipicamente sigmóide padrão (Eq.1).

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

Existem muitos algoritmos de aprendizado específicos para determinado modelos de redes neurais. Neste trabalho a MLP foi treinada usando o algoritmo de *backpropagation* e o termo *momentum*. Mais detalhes sobre ANNs são encontrados em (Haykin, 1998; Haykin, 2001). Neste trabalho o modelo treinado com a rede MLP é simbolizado com ANN2.

Software e metodologia utilizada.

O Waikato Environment for Knowledge Analysis (WEKA) é um conjunto de algoritmos de aprendizado de máquinas para tarefas de mineração de dados, disponível em: <http://www.cs.waikato.ac.nz/~ml/weka/> (Hall et al., 2009). Fornecendo ao usuário uma programação em Java, o WEKA contém ferramentas para pré-processamento de dados, classificação, regressão, regras de associação e visualização. A MLP foi treinada através da ferramenta computacional WEKA. A Figura 4 apresenta os parâmetros utilizados nos modelos de ANNs. Estes parâmetros podem ser alterados, conforme a necessidade da modelagem. As correlações estabelecidas para os modelos de ANNs têm a mesma variável de entrada dos modelos estatísticos.

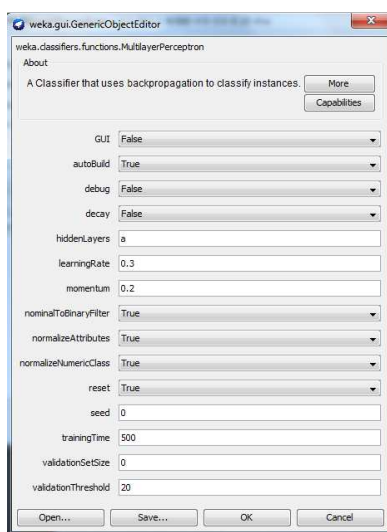


Figura 4: Parâmetros de ANN utilizados na modelagem.

Cálculo da irradiação solar no topo da atmosfera (H_0), fotoperíodo (N) e irradiação direta no topo da atmosfera diária e H_{sc}^d .

A irradiação global no topo da atmosfera (H_0) foi calculada de acordo com equações descritas em (Souza et al., 2005), conforme Eq. (2):

$$H_0 = \frac{24 \times 3600 \times I_0}{\pi} \left[1 + 0,033 \cos \left(\frac{360DJ}{365} \right) \right] \times \left[\left(\frac{\pi}{180^\circ} \omega_s \sin \varphi \sin \delta \right) + (\cos \varphi \cos \delta \sin \omega_s) \right] \quad (2)$$

em que I_0 é a constante solar ($= 1367 \text{ Wm}^{-2}$), DJ é o dia juliano começando em 1 de janeiro (para primeiro de janeiro, $DJ = 1$, e para 31 de dezembro, $DJ = 365$), φ é a latitude local (em graus), δ é a declinação solar (em graus) Eq. (3), ω_s é o ângulo solar horário (em graus) Eq.(4).

$$\delta = 23,45 \times \sin \left[\frac{360(DJ+284)}{365} \right] \quad (3)$$

$$\omega_s = \cos^{-1}[-\tan(\delta) \times \tan(\varphi)] \quad (4)$$

O fotoperíodo (N) é calculado conforme a Eq.(5):

$$N = \frac{2}{15} \times \omega_s \quad (5)$$

A irradiação direta no topo da atmosfera diária é obtida multiplicando a constante solar integrada ($H_{sc} = 4,921 \text{ MJm}^{-2}$) pelo fotoperíodo (Eq. 6):

$$H_{sc}^d = 4,921 \times N \quad (6)$$

Base de dados para validação.

A série climática de 13 anos (1996-2008) foi utilizada para treinamento e validação da rede neural. Foi criado um programa para a separação dos dados. O programa faz a leitura dos cinco primeiros valores, separa os quatro primeiros para o treinamento e o último para validação. Em seguida o programa faz a leitura dos cinco valores sequenciais e executa o mesmo procedimento anterior. Por fim, 80% da série foram utilizadas para treinamento e 20% para validação.

Avaliação estatística.

Na avaliação do desempenho estatístico dos modelos foram utilizados: Mean Bias Error (MBE), Relative Mean Bias Error (rMBE), Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE) e d de Willmott (Santos et al., 2014b):

$$MBE = \frac{\sum_{i=1}^{N'} (P_i - O_i)}{N'} \quad (7)$$

$$rMBE (\%) = 100 \times \frac{\sum_{i=1}^{N'} (P_i - O_i)}{\bar{X}} \quad (8)$$

$$RMSE = \left[\frac{\sum_{i=1}^{N'} (P_i - O_i)^2}{N'} \right]^{\frac{1}{2}} \quad (9)$$

$$rRMSE (\%) = 100 \times \frac{\left[\frac{\sum_{i=1}^{N'} (P_i - O_i)^2}{N'} \right]^{\frac{1}{2}}}{\bar{X}} \quad (10)$$

$$d = 1 - \frac{\sum_{i=1}^{N'} (P_i - O_i)^2}{\sum_{i=1}^{N'} (|P_i| + |O_i|)^2} \quad (11)$$

em que: P_i representa os valores estimados de H_b , O_i os valores medidos de H_b , $|P_i|$ o valor absoluto da diferença $P_i - \bar{O}_1$, em que \bar{O}_1 , representa a média de O_i , $|O_i|$ representa o valor absoluto da diferença $O_i - \bar{O}_1$. \bar{X} é o valor médio medido e N' o número de observações. Valores positivos (negativo) de MBE (rMBE) indicam superestimativa (subestimativa). O RMSE (rRMSE) dá precisão a curto prazo das estimativas com relação aos dados medidos. Os valores de MBE e RMSE são expressos nas mesmas unidades usadas nos dados originais. Os valores de rMBE e rRMSE são apresentados em porcentagem. Diferentes intervalos de rRMSE são definidos para avaliar a acurácia dos modelos (Jamieson et al., 1991; Heinemann et al., 2012):

Excelente se (excellent if) $rRMSE < 10\%$;
 Bom se (good if) $10\% \leq rRMSE < 20\%$;
 Aceitável se (fair if) $20\% \leq rRMSE < 30\%$;
 Pobre se (poor if) $rRMSE \geq 30\%$.

O índice de ajustamento “d” varia de 0 a 1 e representa a total discordância ou a concordância perfeita, respectivamente, entre valores medidos e estimados (Willmott, 1982). A aplicação conjunta de MBE, RMSE e “d” na avaliação dos modelos possibilita uma análise ampla do desempenho dos modelos.

RESULTADOS E DISCUSSÃO

Este trabalho tem como objetivo principal verificar as ANNs na estimativa de H_b diário. Para obtenção de H_b os valores kt_b estimados são multiplicados pela irradiação direta no topo da atmosfera diária (H_{sc}^d).

Estimativas com modelos baseados na transmissividade atmosférica (kt) diária.

Os valores dos índices estatísticos (MBE, rMBE, RMSE, rRMSE e d) de validação dos modelos são mostrados nas Figuras 5 (a, b, c, d, e). Os resultados mostram que o desempenho do modelo ANN2: $MBE = -0,474 \text{ MJm}^{-2}$, $RMSE = 2,701 \text{ MJm}^{-2}$ ($rRMSE = 16,519\%$) e $d = 0,980$, é superior aos modelos estatísticos. O modelo ME4 tem melhor desempenho que os modelos ME5 e ME6. O ME5 tem espalhamento ($rRMSE = 22,662\%$) e alta concordância ($d = 0,961$), mas a subestimativa ($rMBE = -9,364\%$) é superior aos modelos ANN2 e ME4. Este resultado é explicado pelo ajuste do ME4 com dados locais, que favoreceu a superioridade no desempenho.

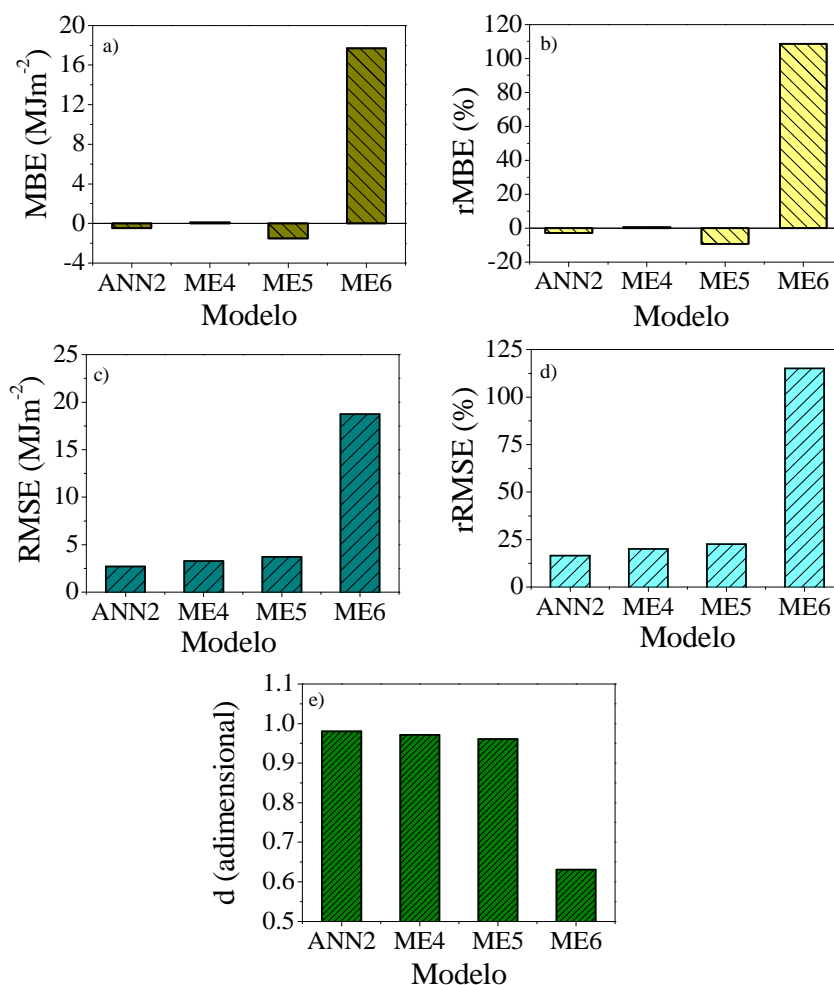


Figura 5: Índices estatísticos de validação dos modelos. a) MBE, b) rMBE, c) RMSE, d) rRMSE, e) d.

Os modelos ANN2 e ME4 tiveram bons desempenhos, com destaque para o modelo de aprendizado de máquinas. Na modelagem diária o ANN2 é indicado como a primeira escolha, seguido do modelo ME4. O modelo de redes neurais estima com uma precisão igual ou superior aos modelos estatísticos. Esta observação comprova que os modelos gerados aqui são boas metodologias na estimativa de H_b^d . A validação dos modelos com a base de dados selecionada da série, através da metodologia escolhida é satisfatória e não comprometeu os resultados finais. Os resultados obtidos neste trabalho são melhores que os encontrados por, Polo et al., (2011), que propuseram um modelo para geração de dados de H_b a cada 10 min e obtiveram rRMSE de 31%. As usinas solares necessitam de dados de H_b em intervalo de tempo menor. Quando é considerada estimativa em partições de tempo menor, os erros encontrados são maiores.

A dispersão do modelo ANN2 e dos modelos estatísticos ME4, ME5 e ME6 para estimativa de H_b^d é mostrada nas Figs. 6 (a, b), respectivamente. Observa-se que a dispersão é similar entre os modelos ANN2, ME4 e ME5. A estimativa com o modelo ANN2 tem melhor ajuste sobre linha de comparação que dos demais modelos, confirmando o menor espalhamento encontrado. A maioria dos pontos com o modelo ANN2 está abaixo da linha ideal de comparação.

O modelo ME6 estima valores elevados de H_b^d com amplitude de ($63,482 \text{ MJm}^{-2}$), diferença entre o mínimo ($3,554 \text{ MJm}^{-2}$) e o máximo ($67,036 \text{ MJm}^{-2}$). Este resultado concorda com índices estatísticos e está relacionado à condição do modelo em estimar resultados mais precisos nos locais com as mesmas características onde foi gerado. O resultado comprova a necessidade de ajustar e validar os modelos nos locais onde serão aplicados. Os coeficientes empíricos de cada modelo podem ser representativos do clima na qual foram ajustados.

Os coeficientes do ME6 foram os responsáveis pelos altos valores de H_b^d estimados. Este modelo foi ajustado num lugar com influência de aerossóis de origem vulcânica (Noroeste do Pacífico), conforme é relatado no artigo original. Com base nas curvas dispersão pode-se destacar que os modelos ANN2 e ME4 ser utilizados para estimar H_b^d com boa precisão. A técnica de aprendizado de máquinas tem melhor capacidade de adequação em todas as condições de clima e cobertura de nuvens.

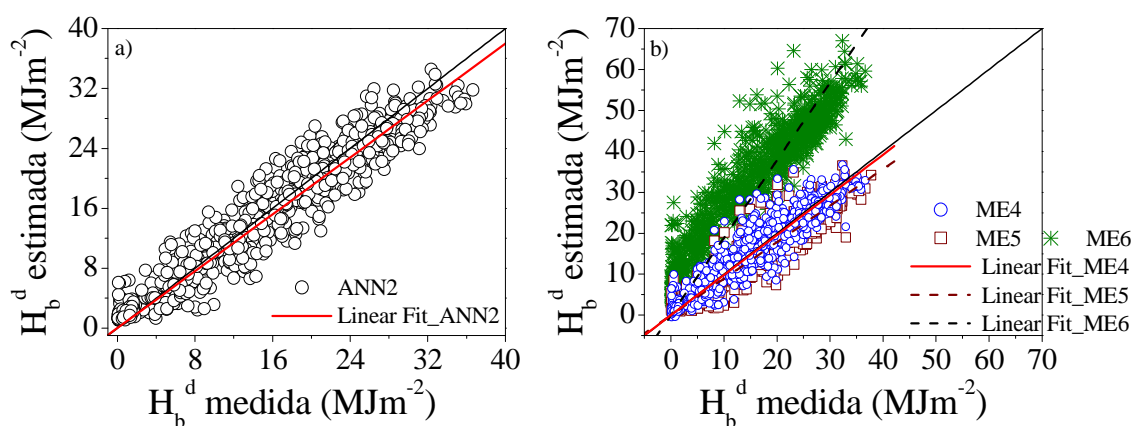


Figura 6: Dispersão entre os valores medidos e estimados dos modelos diários de H_b^d obtidos a partir da relação entre kt_b x kt , a) ANN2 e b) ME4, ME5 e ME6.

O coeficiente de correlação para os modelos diários (Tabela 3) é elevado e variou entre 0,938 e 0,965. É observado que o melhor desempenho ($r = 96,50\%$) é para o modelo ANN2, comprovando o potencial da rede neural em modelar H_b^d . É importante ressaltar que nem sempre elevado coeficiente de correlação indica boa precisão do modelo em estimar determinada variável de interesse (Willmontt, 1982). O modelo ME6 mostra elevada correlação, mas os valores de rMBE, rRMSE e d mostram baixo desempenho do modelo em Botucatu. Por isso, é necessária a utilização de diferentes índices estatísticos de validação de modelos.

Modelo	Regressão	R ²	r
ANN2	$H_{bE}^d = 0,949 * H_{bM}^d$	0,931	0,965
ME4	$H_{bE}^d = 0,979 * H_{bM}^d$	0,891	0,944
ME5	$H_{bE}^d = 0,891 * H_{bM}^d$	0,885	0,941
ME6	$H_{bE}^d = 1,891 * H_{bM}^d$	0,879	0,938

Tabela 3: Regressão dos modelos diários obtidos da Figura 5 (a, b). H_{bE}^d é a irradiação direta na incidência normal estimada; H_{bM}^d é a irradiação direta na incidência normal medida.

Analisar o desempenho da rede neural nas diferentes condições de coberturas de nuvens é interessante na verificação da relação entre as condições de céu predominantes e os erros gerados na estimativa. Nos resultados locais as nuvens são responsáveis pela atenuação de H_b , em sua ausência os aerossóis e o vapor d'água são os principais fatores atenuantes.

Estimativa de H_b com modelos de redes neurais: Estudo de caso para condição de céu sem nuvens.

A estimativa na condição de céu sem nuvens é realizada na verificação do efeito dos aerossóis e do vapor d'água na modelagem de H_b . Os dados utilizados foram para $kt > 0,65$. Foi utilizado o Percentage Split do programa WEKA, com 70% dos dados utilizados para treinamento e 30% para validação. A escolha dos dados para treinar e validar é feita aleatoriamente, sem tendência a série na separação. Na condição de céu sem nuvens o ANN2 superestimou e tem espalhamento menor que na validação com todas as condições de cobertura de nuvens (Figura 7). O ANN2 teve $rRMSE = 12,813\%$. Climaticamente, o resultado mostra que as nuvens são responsáveis por aumentarem consideravelmente a dispersão de H_b .

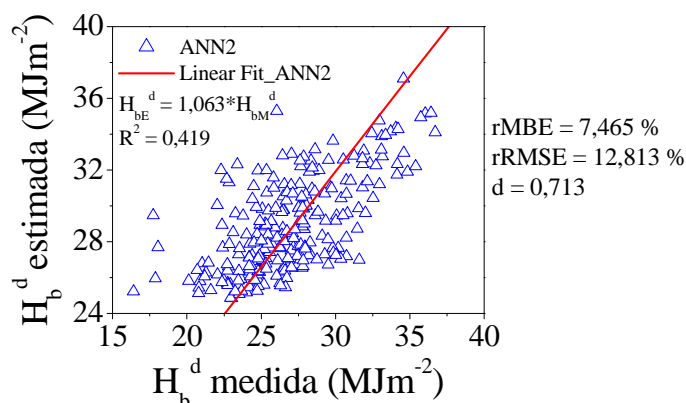


Figura 7: Dispersão entre os valores medidos e estimados para condição de céu claro e índices estatísticos de validação.

Na ausência de nuvens, a variabilidade dos valores de $rRMSE$ e d está relacionada com a flexibilidade da rede neural em estimar H_b e provavelmente ao efeito do espalhamento causado pelos aerossóis e absorção do vapor d'água. A dispersão para condição de céu sem nuvens é caracterizada por ser menor que na condição de todos os tipos de cobertura de céu. Este resultado é quantificado pelos valores de $rMBE$ e $rRMSE$ e qualificado pela curva de dispersão. Quando são observados valores baixos de H_b associados com estimativas mais elevadas, a mesma explicação para a variabilidade de $rRMSE$ também é válida aqui.

CONCLUSÕES

As dispersões observadas são causadas pela nebulosidade. Na condição de céu sem nuvens a dispersão é causada, principalmente, pela atenuação da radiação pelos aerossóis e pela absorção por vapor d'água.

O modelo neural utilizado teve desempenho superior aos modelos estatísticos e deve ser a primeira opção na modelagem de H_b . O modelo estatístico com ajuste local tem desempenho melhor que os demais modelos estatísticos avaliados. Isto mostra a necessidade do ajuste do modelo com dados locais. A estimativa de H_b foi satisfatória e pode ser empregada com sucesso nos locais com escassez

de informações solarimétrica. O WEKA mostrou ser uma ferramenta computacional que pode ser utilizada na estimativa da H_b .

Em trabalhos futuros, realizar-se-á estudo comparativo entre a rede neural com modelos físicos de transferência radiativa na estimativa de H_b . A rede pode ser utilizada para avaliar o potencial na modelagem de outras radiações. O desenvolvimento de novos modelos de redes neurais para estimativa de H_b em partições de tempo menores (10 e 5 min, por exemplo) é o passo seguinte da pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Ångström, A. K. (1924) Solar and terrestrial radiation. *Quarterly Journal of the Royal Meteorological Society*, v.50, p.121.
- Escobedo, J. F.; Gomes, E. N.; Oliveira, A. P.; Soares, J. (2011). Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. *Renewable Energy*, v.36, p.169-178.
- Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. (2009). *The WEKA Data Mining Software: An Update; SIGKDD Explorations*, v.11, Issue 1.
- Hasni, A.; Sehli, A.; Draoui, B.; Bassou, A.; Amieur, B. (2012). Estimating global solar radiation using artificial neural network and climate data in the south-western region of Algeria. *Energy Procedia*, v.18, p.531 – 537.
- Haykin, S. (1998). *Neural networks: A comprehensive foundation*. 2nd ed. Hamilton: Prentice Hall, 897p.
- Haykin, S. (2001). *Redes Neurais: Princípios e prática*. Trad. Paulo Martins Engel. – 2. Ed. – Porto Alegre: Bookman, 900p.
- Heinemann, A. B.; Van Oort, P. A. J.; Fernandes, D. S.; Maia, A. H. N. (2012). Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. *Bragantia*, Campinas, v. 71, n. 4, p.572-582.
- 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, p.337-350.
- Jiang, Y. (2009). Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models. *Energy*, v.34, p.1276-1283.
- Kalogirou, S. A.; Mathioulakis, E.; Belessiotis, V. (2014). Artificial neural networks for the performance prediction of large solar systems. *Renewable Energy*, v.63, p.90-97.
- Karatepe, E.; Boztepe, M.; Colak, M. (2006). Neural network based solar cell model. *Energy Conversion and Management*, v.47, p.1159-1178.
- Koca, A.; Oztop, H. F.; Varol, Y.; Koca, G. O. (2011). Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. *Expert Systems with Applications*, v.38, p.8756-8762.
- Liu, B. Y. H.; Jordan, R. C. (1960). The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy*, v.3, p.1-19.
- Lopez, G.; Rubio, M. A.; Batlles, F. J. (2000). Estimation of hourly direct normal from measured global solar irradiance in Spain. *Renewable Energy*, v.21, p. 175-186.
- Lorena, A. C.; Jacintho, L. F. O.; Siqueira, M. F.; Giovann, R.; Lohmann, L. G.; Carvalho, A. C.P.L.F.; Yamamoto, M. (2011). Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*, v.38, p.5268–5275.
- Louche, A.; Notton, G.; Poggi, P.; Simonnot, G. (1991). Correlations for direct normal and global horizontal irradiation on a french mediterranean site. *Solar Energy*, v.46, p. 261-266.
- Lyra, G. B.; Zanetti, S. S.; Santos, A. A. R.; Souza, J. L.; Lyra, G. B.; Oliveira-Júnior, J. F.; Lemes, M. A. M. (2015). Estimation of monthly global solar irradiation using the Hargreaves–Samani model and an artificial neural network for the state of Alagoas in northeastern Brazil. *Theoretical and Applied Climatology*. DOI 10.1007/s00704-015-1541-8.
- Nfaoui, H.; Buret, J. (1993). Estimation of daily and monthly direct, diffuse and global solar radiation in Rabat (Morocco). *Renewable Energy*, v. 3, p. 923-930.
- Oliveira, A. L. I. (2006). Estimation of software project effort with support vector regression. *Neurocomputing*, v.69, p. 1749-1753.

- Ozgoren, M.; Bilgili, M.; Sahin, B. (2012). Estimation of global solar radiation using ANN over Turkey. *Expert Systems with Applications*, v.39, p.5043-5051.
- Polo, J.; Zarzalejo, L. F.; Marchante, R.; Navarro, A. A. (2011). A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. *Solar Energy*, v.85, p.1164–1170.
- Qazi, A.; Fayaz, H.; Wadi, A.; Raj, R. G.; Rahim, N. A.; Khan, W. A. (2015). The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. *Journal of Cleaner Production*, v.104, p.1-12.
- Reddy, K. S. (2003). Solar resource estimation using artificial neural networks and comparison with other correlation models. *Energy Conversion and Management*, v.44, p.2519–2530.
- Santos, C. M, Souza, J. L, Ferreira Junior, R. A.; Tiba, C.; Melo, R. O.; Lyra, G. B.; Teodoro, I.; Bastos Lyra, G.; Lemes, M. A. M. (2014a). On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. *Energy*, v.71, p.388-398.
- Santos, C. M.; Souza, J. L.; Teramoto, E. T.; Tiba, C.; Melo, R. O. (2014b). Modelagem da irradiação solar global média horária mensal (H_g^h) para quatro localidades de Alagoas/Brasil. *Nativa, Sinop*, v.2, p. 79-88.
- 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*, v.79 201–214.
- Souza, J. L.; Nicácio, R. M.; Moura, M. A. L. (2005). Global solar radiation measurements in Maceió, Brazil. *Renewable Energy*, v. 30, p. 1203 – 1220.
- Vignola, F.; Mcdaniels, D. K. (1986). Beam-global correlations in the Pacific Northwest. *Solar Energy*, v. 36, p. 409-418.
- Wana, K. K. W.; Tanga, H. L.; Yang, L.; Lam, J. C. (2008). An analysis of thermal and solar zone radiation models using an Angstrom–Prescott equation and artificial neural networks. *Energy*, v.33, p.1115–1127.
- Willmott, C. J. (1982). On the validation of models. *Physical Geography, Delaware*, v. 2, p. 184-194.
- WMO – World Meteorological Organization. (2008). Guide to meteorological Instruments and Methods of Observation. WMO-n°8, Seventh Edition, p.1-681, Geneva, Switzerland.
- Yadav, A. K.; Chandel, S. S. (2014). Solar radiation prediction using Artificial Neural Network techniques: A review. *Renewable and Sustainable Energy Reviews*, v.33, p.772–78.
- Zarzalejo, L. F.; Ramirez, L.; Polo, J. (2005). Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index. *Energy*, v.30, p.1685–1697.

ESTIMATION OF BEAM AT NORMAL INCIDÊNCIA USING ARTIFICIAL NEURAL NETWORK AND COMPARING WITH STATISTICAL MODELS

ABSTRACT: The direct irradiation at normal incidence (H_b) was estimated using Artificial Neural Networks (ANN2) to a database from 1996 to 2008 in Botucatu-Brazil, region with influence of agricultural burning. The network used was the Multilayer Perceptron (MLP), with the *backpropagation* algorithm and *momentum* term. The ANN2 was compared with a statistical model of local adjustment (ME4) and two models selected in the literature (ME5 and ME6). The input variable was the k_t (atmospheric transmissivity of global radiation). The models were evaluated by means of statistical indicatives: Mean Bias Error (MBE, MJm^{-2}), Relative Mean Bias Error (rMBE, %), Root Mean Square Error (RMSE, MJm^{-2}), Relative Root Mean Square Error (rRMSE, %) and d Willmott. The ANN2 performance is better than statistical models. The estimate with ANN2 resulted: rRMSE=16.519% and d=0.980. The ME4 estimated with rRMSE=20.130% and d=0.971. The ME5 and ME6 had underperformed ME4. The ANN2 model is recommended as the first choice to estimate H_b , followed by ME4 model.

Keywords: Solar Radiation, WEKA, Data Mining, Learning Machines, Modelling.