

AVALIAÇÃO DO ALGORITMO SEQUENTIAL MINIMAL OPTIMIZATION (SMO) NA ESTIMATIVA DA FRAÇÃO TRANSMITIDA DA IRRADIAÇÃO DIRETA NA INCIDÊNCIA NORMAL (kt_b^d)

Cícero Manoel dos Santos – ciceromanoel2007@gmail.com

João Francisco Escobedo – escobedo@fca.unesp.br

Maurício Bruno Prado da Silva – mauricio.prado19@hotmail.com

Universidade Estadual Paulista, FCA-Botucatu, Departamento de Engenharia Rural

Silvia Helena Modenese Gorla da Silva – silvia@registro.unesp.br

Érico Tadao Teramoto – eteramoto@registro.unesp.br

Universidade Estadual Paulista, Campus de Registro, Departamento de Eng^a. de Pesca

Resumo. Neste trabalho o algoritmo Sequential Minimal Optimization (SMO) é utilizado para estimar a fração transmitida da irradiação direta na incidência normal diária (kt_b^d). A Radial Basis Function (RBF) de Kernel é empregada para regressão. O modelo estatístico (#M3) é desenvolvido e comparado com o modelo (SMO3). A variável de entrada utilizada é a razão de insolação (n/N) [n é a duração do brilho solar e N o fotoperíodo]. Foram utilizadas medidas de 13 anos para região de Botucatu – SP. Duas bases de dados denominadas de ano típico (AT) e ano atípico (AAT), selecionados da base total de 13 anos, são utilizados para validação dos modelos. Na avaliação dos modelos foram usados: Relative Mean Bias Error (rMBE), Relative Root Mean Square Error (rRMSE), erro percentual relativo (ϵ) e d de Willmott. O SMO3 tem precisão melhor que o modelo #M3. A validação com AT e AAT foi satisfatória. Por último, o desempenho da SMO3 na estação chuvosa e seca é analisado para verificar a influência das nuvens, aerossóis e vapor d'água na dispersão das estimativas e o aumento dos erros. O SMO estima kt_b^d com melhor precisão e pode ser utilizado.

Palavras-chave: SVM, razão de insolação, WEKA.

1. INTRODUÇÃO

A irradiação global (H_G), irradiação difusa (H_D) e a irradiação direta na incidência normal (H_b) possuem grande importância em estudos climáticos, agrônômicos, na engenharia e como fonte de energia renovável. A H_G é monitorada por muitos centros de pesquisas, no entanto, poucos são os que monitoram rotineiramente H_D e mais escassas são as medidas de H_b (Kotti et al., 2014). O conhecimento da variabilidade espacial e temporal de H_b é indispensável na concepção e dimensionamento de projetos para geração de energia solar limpa e renovável, incluindo fotovoltaica, térmica e as tecnologias concentradas (Blanc et al., 2015). Nos últimos anos têm-se buscado meios de se obter H_b nos locais onde não existem estações solarimétricas para geração de curvas anuais ou de mapas solares.

Nos locais onde não existem estações solarimétricas com sensores monitorando H_b , a modelagem torna-se uma alternativa na estimativa e avaliação dos recursos solares. Várias metodologias têm sido utilizadas para estimar H_b . A relação proposta por Ångström (1924) tem sido utilizada para correlacionar kt_b com a razão de insolação (n/N). Fernandez-Peruchena et al., (2014) desenvolveram um método para estimativa horária (H_b^h) a partir de medidas diárias. Padovan et al., (2014) estimaram H_b com diferentes modelos estatísticos para instalações de sistemas de concentradores solares e obtiveram resultados com boa precisão. Modelos de transferência radiativa, a partir de satélites, têm sido desenvolvidos e utilizados, no entanto, a resolução não é suficiente para avaliar sistemas solares (Beyer et al., 2010; Viana et al., 2011).

Nos últimos anos, Máquina de Vetor de Suporte (SVM) têm ganhado grande destaque na modelagem solar. A SVM tem sido utilizada numa série de estudos para estimar a H_G e tem mostrado ser mais eficiente que os modelos empíricos (Chen et al., 2013; Chen e Li, 2014). As técnicas de aprendizado de máquinas (AMs) estão sendo disponibilizadas em Softwares. Dentre estes, destaca-se o Waikato Environment for Knowledge Analysis (WEKA), que tem seu conjunto de algoritmos de AMs amplamente utilizados com sucesso em diversas áreas (Lorena et al., 2011; Yadav e Chandel, 2015). O algoritmo Sequential Minimal Optimization (SMO) é destaque por resolver problemas com SVM utilizando programação quadrática, com potencial para aplicação nas diversas áreas (Platt, 1998). Apesar do sucesso da aplicação em determinadas áreas, o SMO é um algoritmo de regressão que não tem sido avaliado na modelagem solar e, tão pouco, nos problemas de predição de H_b .

Neste sentido, mostra-se no trabalho que SMO pode ser aplicada com sucesso na estimativa de kt_b a partir de medidas de n . A relação neste trabalho é de kt_b vs n/N (diário). A estimativa através de kt_b elimina-se o efeito astronômico nas variáveis e evidencia somente o climático. O desempenho de SMO é comparado com o desempenho de um modelo estatístico de mesma variável de entrada, ajustados com dados locais. Este trabalho também discute a validação destes modelos com duas bases de dados, aqui definidas de ano típico (AT) e ano atípico (AAT). O trabalho esta organizado da seguinte forma: na metodologia são mostrados os dados utilizados, o local, descrição da SVM, descrição do algoritmo SMO, a geração do modelo estatístico, descrição da obtenção da base de validação e os critérios

de validação. O desempenho dos modelos é apresentado nos resultados e discussão. Um estudo de caso é realizado para verificar o desempenho do SMO em duas estações distintas: chuvosa e seca, destacando o efeito das nuvens, aerossóis e vapor d'água. O artigo é finalizado com as conclusões mais relevantes.

2. MATERIAL E MÉTODOS

2.1 Local de estudo e dados

Neste trabalho, os dados utilizados foram da cidade de Botucatu, localizada no Estado de São Paulo – Brasil. Botucatu é a sexta cidade mais populosa da região Centro-Oeste Paulista (138.000 habitantes) e possui área total de 1.483 m². A Estação de Radiometria solar está situada na Faculdade de Ciências Agrônomicas da UNESP/Botucatu (latitude 22,85°S; longitude 48,45°W e altitude 786m). A Fig. 1 ilustra a localização de Botucatu e da estação radiométrica. Botucatu é a única cidade da região Centro-Oeste Paulista que possui medidas confiáveis de H_b de longo prazo. Devido a problemas técnicos no pireliômetro, a estação parou de monitorar H_b em 2009. Mas continua monitorando rotineiramente outras radiações (H_G e H_D , por exemplo). Assim, H_b pode ser obtida indiretamente da diferença entre H_G e H_D . Neste sentido, a estimativa de H_b usando novas técnicas é de grande importância para ampliação da série histórica e recuperação de dados.

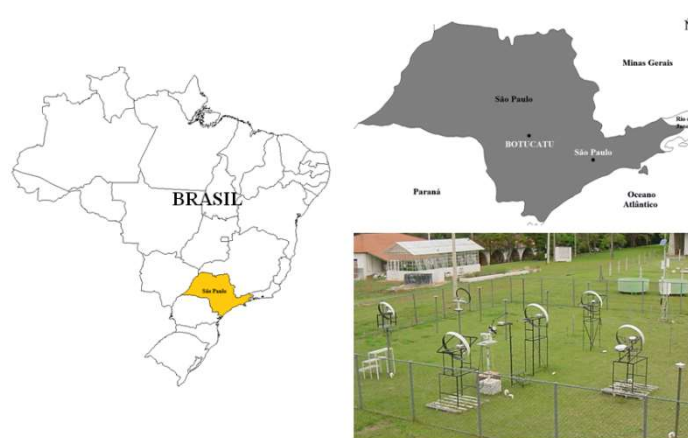


Figura 1 - Localização da estação radiométrica.

Com bioma de cerrado e mata atlântica, Botucatu possui clima temperado quente (mesotérmico), verão quente e úmido com elevada precipitação e inverno seco (Escobedo et al., 2011). 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 do Brasil.

Para uma modelagem mais precisa, o uso de uma longa série climática é de grande importância na área solar. Neste estudo, foram utilizadas dadas diários de irradiação direta na incidência normal (H_b) e duração do brilho solar (n), do período de fevereiro de 1996 a dezembro de 2008. A H_b foi obtida por pireliômetro da Eppley NIP acoplado a um rastreador solar ST3 da Eppley e n foi obtido por um heliógrafo convencionais do tipo Campbell-Stokes.

2.2 Máquina de Vetor de Suporte (SVM)

SVM é um método de aprendizagem de máquina supervisionado derivado da teoria de aprendizagem estatística, que analisa bancos de dados e reconhece padrões. Este método é baseado no princípio da minimização do risco estrutural, que tenta minimizar um limite superior de erro de generalização em vez de minimizar o erro de treinamento local (Chen et al., 2015). O princípio fundamental é fazer a correlação por meio do mapeamento não-linear dos conjuntos de dados de entradas para um espaço de alta dimensão, usando um conjunto de funções de Kernel. Maiores detalhes da teoria da SVM podem ser encontrados em (Vapnik, 1995).

Neste trabalho o algoritmo Sequential Minimal Optimization (SMO) é utilizado para regressão. SMO é um algoritmo simples que rapidamente soluciona e resolve problema de SVM com programação quadrática de menor otimização possível com dois multiplicadores de Lagrange, maiores detalhes são encontrados em (Smola e Schölkopf, 2004).

Pela eficiência computacional, simplicidade e adaptação para otimização de problemas mais complexos, neste estudo a função de Kernel Radial Basis Function (RBF) é a escolhida para regressão com SMO. Na utilização da RBF é necessário ajustar adequadamente seus parâmetros C (custo), γ (gama) e ϵ (épsilon). Os valores de C e γ são testados e aqueles com a melhor precisão na validação-cruzada são selecionados. O valor de ϵ controla como a função vai ajustar os dados de treinamento. O algoritmo SMO e a função de Kernel foram aplicados utilizando a ferramenta computacional Waikato Environment for Knowledge Analysis (WEKA).

O WEKA é um conjunto de algoritmos de AMs para tarefas de mineração de dados, disponível em: <<http://www.cs.waikato.ac.nz/~ml/weka/>> (Hall et al., 2009). Foi estabelecida correlação entre a fração transmitida da irradiação direta na incidência normal (kt_b^d) com a razão de insolação (n/N)^d (Tab. 1). Maiores detalhes sobre mineração de dados e o uso WEKA nas diferentes modelagens são encontrados em (Witten et al., 2011).

Tabela 1 - Modelo SMO3.

Partição	Modelo	Relação
Diário	SMO3	kt_b^d vs $(n/N)^d$

2.3 Modelo Estatístico

O modelo estatístico é desenvolvido conforme metodologia de Ångström (1924). A fração transmitida diária da irradiação direta na incidência normal kt_b^d foi relacionada com a razão de insolação ($r = n/N$), similarmente ao modelo SMO3. A Fig. 2 apresenta a correlação entre kt_b^d e $(n/N)^d$. Na geração da equação para esse modelo foi utilizado 4.042 dias. As maiores dispersões são para valores de $(n/N)^d > 0,50$ e um único valor de $(n/N)^d$ pode corresponder a diferentes valores de kt_b^d . O modelo foi gerado através de ajuste polinomial do 2º grau. Funções de grau superior não resultaram em melhorias no ajuste do modelo.

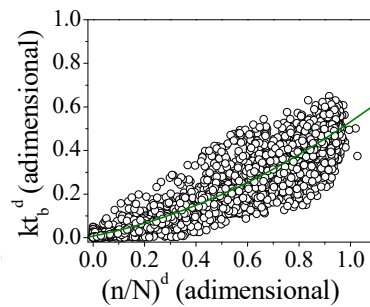


Figura 2 - Correlação para geração do modelo entre kt_b^d e $(n/N)^d$.

O modelo estatístico (#M3) gerado e seu respectivo coeficiente de determinação (R^2):

$$kt_b^d = 0,007 + 0,239 \times (n/N) + 0,285 \times (n/N)^2 \quad (R^2 = 0,74) \quad (1)$$

2.4 Base de dados para o treinamento e validação de SMO3 e de #M3

Neste estudo a série climática de 13 anos foi utilizada no treinamento do SMO3. Como não existe uma regra geral para a escolha da base de dados para validação. Aqui foi utilizada uma base de dados das irradiações (horária e diária) denominada por ano típico (AT) e ano atípico (AAT), os quais foram selecionados da série. A seleção do AT e AAT, foi através de análises estatísticas onde foi comparado para cada mês do ano, o valor da irradiação média horária inter-anos considerando o intervalo de variação do desvio padrão da média com o valor da irradiação média de cada ano. A seleção horária foi estendida para os dados diários. A Fig. 3 apresenta o exemplo da seleção dos meses de fevereiro e agosto. As barras na vertical representam a irradiação horária média de cada ano, a linha horizontal contínua representa a irradiação média inter-anos (média dos valores médios dos 13 anos de dados) com o desvio padrão.

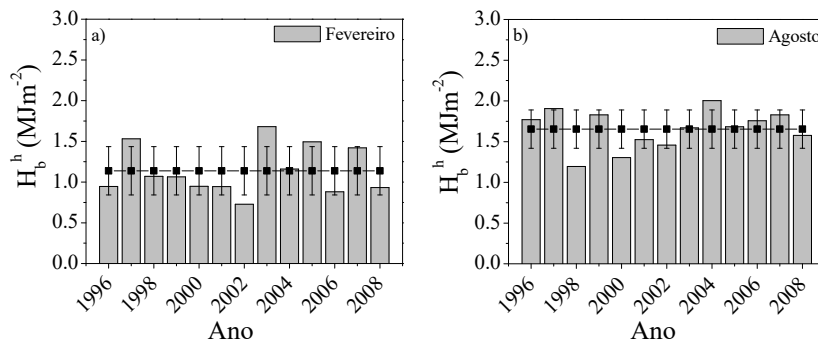


Figura 3 - Comparação do valor da irradiação média inter - anos com desvio padrão e o valor da irradiação média do ano, para os meses: a) Fevereiro e b) Agosto.

O processo de seleção do AT é similar ao processo de seleção do Ano Meteorológico Típico (WMO, 1981). Para o mês de fevereiro, o AT foi 2004, enquanto que em agosto foi 2003. Nestes dois meses, os valores das irradiações médias inter-anos praticamente coincidiram com os valores das irradiações médias do ano. Ao contrário, no AAT, os valores das irradiações médias do ano estão mais distantes (para acima ou para baixo) dos valores das irradiações médias inter-anos. Assim para o mês de fevereiro, o AAT foi 2003, enquanto que para agosto foi 1998. A constituição da base dos meses para o AT e AAT é representada na Tab. 2.

Tabela 2 - Base que constituem o AT e AAT.

		Meses											
		J	F	M	A	M	J	J	A	S	O	N	D
AT	2004	2004	1998	2006	2007	1998	2000	2003	1996	1997	2006	2004	
ATT	1998	2003	2000	2008	2003	2003	2008	1998	2003	2003	1998	1996	

2.5 Cálculo do fotoperíodo (N) e H_{sc}^d

O fotoperíodo (N, em horas) é calculado conforme a Eq.(2):

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

ω_s é o ângulo solar horário do nascer do sol (em graus) Eq.(3).

$$\omega_s = \cos^{-1}[-\tan(\varphi) \times \tan(\delta)] \quad (3)$$

φ é a latitude local (em graus), δ é a declinação solar (em graus) Eq.(4):

$$\delta = 23,45 \times \sin \left[\frac{360(DJ+284)}{365} \right] \quad (4)$$

DJ é o dia juliano começando em 1 de janeiro (para primeiro de janeiro, DJ = 1, e para 31 de dezembro, DJ = 365). A irradiação direta no topo da atmosfera diária (H_{sc}^d , MJm^{-2}) é obtida multiplicando a constante solar integrada ($H_{sc} = 4,921 MJm^{-2}$) pelo fotoperíodo (N), Eq.(5):

$$H_{sc}^d = 4,921 \times N \quad (5)$$

2.6 Indicadores de validação

O Relative Mean Bias Error (rMBE), Relative Root Mean Square Error (rRMSE), erro percentual relativo (e), d de Willmott e coeficiente de determinação (R^2), foram utilizados para certificar a validade e precisão dos modelos desenvolvidos (Escobedo et al., 2011; Santos et al., 2014). Os índices são obtidos pelas relações:

$$rMBE (\%) = 100 \times \frac{\frac{\sum_{i=1}^x (H_E - H_M)}{x}}{\bar{X}} \quad (6)$$

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

$$e(\%) = \left[\frac{(\bar{H}_M - \bar{H}_E)}{\bar{H}_M} \right] \times 100 \quad (8)$$

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

em que: H_E representa os valores estimados, H_M os valores medidos, $|H'_E|$ o valor absoluto da diferença $H_E - H_M$, em que \bar{H}_M , representa a média de H_M , $|H'_M|$ representa o valor absoluto da diferença $H_M - \bar{H}_M$ e \bar{H}_E a média dos valores estimados. $\bar{X} \left(= \frac{1}{x} \sum_{i=1}^x O_i \right)$ é o valor médio da medida e x o número de observações. Neste trabalho, os seguintes critérios são utilizados para estraficar os 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 R^2 fornece uma medida da relação linear entre os valores medidos e estimados. Os valores de R^2 variam de 0 a 1, e quanto maior seu valor, melhor o ajuste do modelo:

$$R^2 = \frac{\sum_{i=1}^x (\hat{H}_M - \bar{H}_M)^2 - \sum_{i=1}^x (H_E - \bar{H}_M)^2}{\sum_{i=1}^x (H_M - \bar{H}_M)^2} \quad (10)$$

3. RESULTADOS E DISCUSSÃO

Neste trabalho, o algoritmo SMO foi utilizado para estimar a fração transmitida da irradiação direta na incidência normal diária (kt_b^d). O desempenho do SMO é comparado com modelo estatístico (#M3). Os resultados obtidos são desenvolvidos e discutidos, primeiramente, para a dispersão dos valores estimados com base na validação do AT e AAT, em seguida são analisados os valores de rMBE, rRMSE e d. Os resultados mais significativos são mostrados.

Primeiramente foram encontrados os valores dos parâmetros da função RBF de Kernel. Os parâmetros da RBF: C , ε e γ foram ajustados através de inúmeros ensaios com diferentes combinações. Foram considerados os seguintes conjuntos de valores dos parâmetros C , ε e γ : - $C = \{10, 50, 100, 200 \text{ e } 400\}$, $\varepsilon = \{0,001; 0,002; 0,003; 0,004 \text{ e } 0,005\}$ e $\gamma = \{0,1; 0,2; 0,3; 0,4 \text{ e } 0,5\}$. Portanto, para SMO3 com a função RBF de kernel foram realizados 60 simulações, abrangendo as diferentes combinações dos parâmetros. Os valores mais significativos e utilizados foram: $C = 50$, $\varepsilon = 0,005$ e $\gamma = 0,2$. Obteve-se para o modelo SMO3 coeficiente de determinação (R^2) igual 0,87. Este valor é maior que do modelo #ME3 e indica bom ajuste do SMO.

3.1 Análise de precisão do modelo SMO3

Os valores dos índices de validação (rMBE, rRMSE e d) e a curva dispersão do modelo SMO3, validado com AT e ATT, são apresentados na Fig. 4a. A estimativa com o SMO3 mostra que os pontos crescem com a linha ideal de comparação 1:1. No intervalo de kt_b^d entre 0,30 e 0,60, onde a maior parte dos pontos se encontra os valores mais elevados de Δ são encontrados Fig. 4b. O Δ é a diferença entre as medidas e as estimativas. O Δ apresenta comportamento similar a dispersão, com as estimativas subestimando para $kt_b^d > 0,50$, que correspondem a dias de condição de céu de limpo. Na validação do modelo de kt_b^d vs $(n/N)^d$, a base de validação utilizada é satisfatória e pode ser considerada em outras pesquisas.

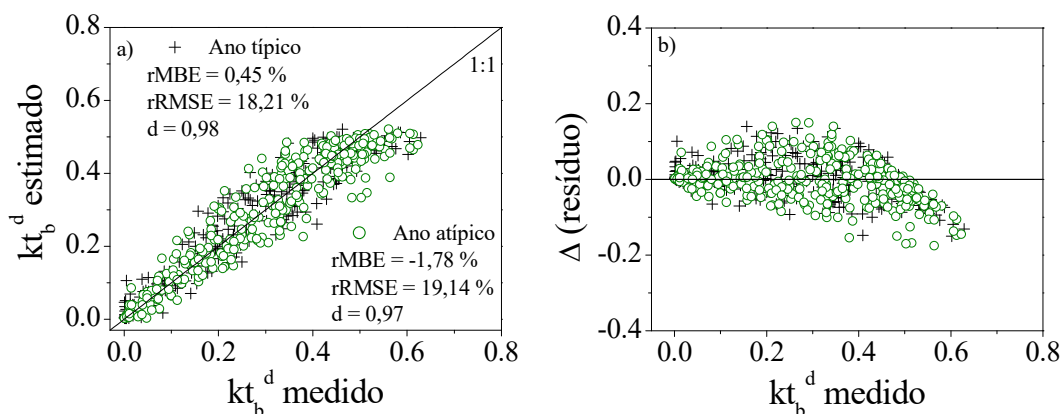


Figura 4 - Dispersão e índices de validação do modelo SMO3 na estimativa de kt_b^d (a) e Δ (b).

Considerando rRMSE, o SMO3 estima kt_b^d com boa precisão. A validação com o AT (*SMO3) tem melhor estimativa: rMBE = 0,45%, rRMSE = 18,21% e d = 0,98, que o ano ATT **SMO3 (rMBE = -1,78%, rRMSE = 19,14% e d = 0,97), respectivamente. Os valores de d são elevados e próximos de 1, indicando que o SMO3 é potencialmente recomendada para estimar kt_b^d . O modelo SMO3 com o AAT subestimou as medidas e com AT superestimou. Problemas na queima do heliograma e erros operacionais na leitura pode afetar a estimativa de kt_b^d a partir de n/N .

3.2 Análise de precisão do modelo #M3

A dispersão e os valores dos índices de validação do modelo #M3 são mostrados na Fig. 5a. A dispersão de kt_b^d medido em função da estimativa é maior para valores de $kt_b^d > 0,4$. O Δ do #ME3 tem maior espalhamento para valores

de $kt_b^d > 0,30$ (Fig. 5b). A validação com **#M3 (ATT) tem $rRMSE = 22,89\%$ e subestimativa ($rMBE = -6,86\%$). O *#M3 (AT) subestimou ($rMBE = -3,45\%$), $rRMSE = 20,79\%$ e $d = 0,96$. Com base nos valores de $rRMSE$, o #M3 estima com precisão aceitável. Por fim, os resultados aqui apresentados são satisfatórios e a curva de dispersão mostra a limitação que o #M3 possui quando comparado com a SMO3. Destaca-se a boa generalização da base de dado utilizada na validação.

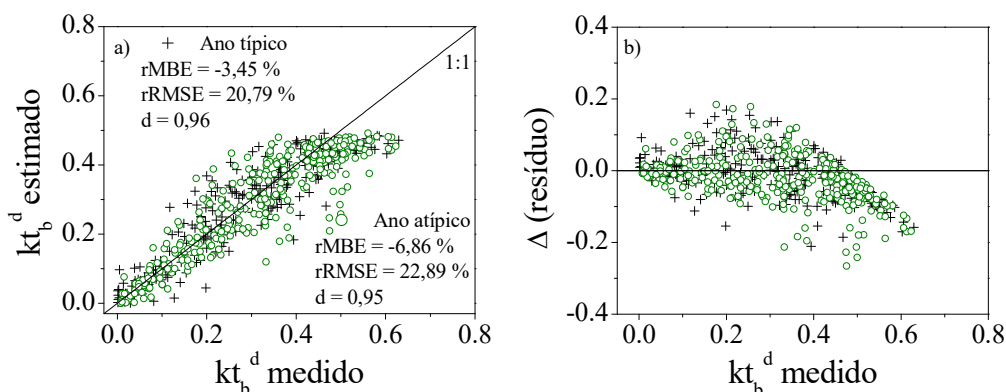


Figura 5 - Dispersão e valores dos índices de validação do modelo #M3 na estimativa de kt_b^d , (a) dispersão e (b) Δ .

3.3 Análise de precisão de SMO na estimativa de kt_b^d nas estações chuvosa e seca: estudo de caso

Nesta seção um estudo de caso para avaliação de SMO3 é desenvolvido. A série de 13 anos é separada em duas: estação chuvosa e estação seca. Do total dos dados utilizados de cada estação, 80% foram para treinamento e 20% na validação. Esta divisão é considerada para analisar o efeito das nuvens, dos aerossóis e do vapor d'água no desempenho da SMO e na dispersão de kt_b^d . Este estudo é motivado pela região de Botucatu ser fortemente afetada pela queima de cana de açúcar para a colheita mecanizada e queimadas florestal, que aumentam significativamente a concentração de materiais particulados na atmosfera local (Codato et al., 2008). A concentração de vapor d'água na atmosfera tem origem das represas hidroelétricas situadas num raio de 70 km de Botucatu e, muito provavelmente, das usinas sucroalcooleira no período da colheita e moagem em regiões adjacentes. Nesta seção o modelo é simbolizado como SMO3.3 (kt_b^d vs n/N diário).

A Fig. 6 mostra os valores dos índices de validação e a dispersão das estimativas para o modelo na estação chuvosa e na estação seca. O SMO3.3 estima com aceitáveis $rRMSE$ nas duas estações. O valor do $rRMSE$ na estação chuvosa (22,06%) é menor que na estação seca (24,67%). Estes valores são maiores que os obtidos na modelagem com SMO3. Os resultados observados com SMO3.3 são justificados pelo aumento das nuvens no período chuvosa, enquanto que na ausência das nuvens os aerossóis e o vapor d'água elevam o espalhamento. Por fim, o SMO mostra ser eficiente para estimativa de kt_b^d nas diferentes condições atmosféricas e tem potencial para aplicação em locais de interesse com resultados satisfatórios.

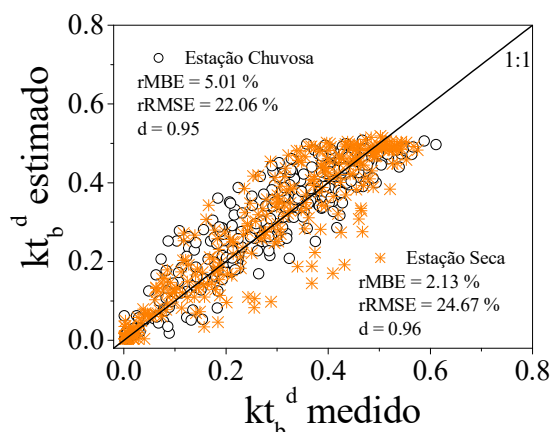


Figura 6 - Dispersão e índices de validação do SMO3.3.

Para verificar a adequação da SMO3 na estimativa de kt_b^d , a precisão é comparada com o modelo estatístico. A comparação é viável uma vez que os parâmetros de entrada são as mesmas. Esta comparação é feita para mostrar qual a metodologia sobre-sai na estimativa de kt_b .

A Fig. 7 mostra o erro percentual relativo "e" obtido na validação dos modelos. Considerando os limites de variação de e ($\pm 10\%$), os resultados obtidos são aceitáveis para os modelos aqui avaliados. Em geral, a validação com

SMO3 tem menor valor de e quando comparado com #M3. SMO3 validado com AT tem o valor de e negativo e positivo quando validado com o AAT. A superioridade das técnicas de aprendizado de máquina com relação aos modelos empíricos (estatísticos) é destacada na literatura (Mohammadi et al., 2015; Chen et al., 2015). Por fim, a comparação pareada dos modelos aqui apresentados mostra que SMO3 tem potencial para estimar kt_b^d .

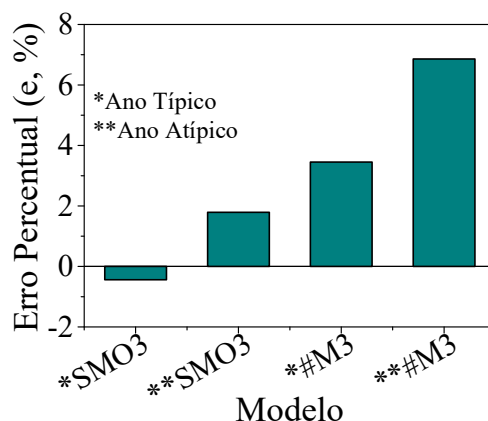


Figura 7 - Erro percentual relativo (e , %) dos modelos SMO3 e #M3.

4. CONCLUSÕES

O trabalho compara o desempenho do algoritmo SMO3 com modelo estatístico na partição diária. O modelo SMO3 apresenta bom ajuste na modelagem, com elevado coeficiente de determinação (R^2). O valor de R^2 gerado pelo modelo SMO3 é maior que do modelo estatístico. Para a função de Kernel RBF os parâmetros utilizados e recomendados na modelagem de kt_b^d são: $C = 50$, $\gamma = 0,2$ e $\varepsilon = 0,005$. A variável de entrada foi capaz de modelar kt_b^d . A base de validação adotada, AT e AAT, teve desempenho satisfatório. Na estação seca os aerossóis e o vapor d'água são responsáveis pela atenuação da radiação e no aumento de rRMSE. Já na estação chuvosa, o aumento no espalhamento é causado pela atenuação da radiação pelas nuvens. A comparação mostra desempenho superior do modelo de SMO com relação ao estatístico. O SMO3 é indicado para estimativa da kt_b como a primeira opção devido sua capacidade de generalização e adequação.

Os resultados mostram que o SMO pode servir como uma alternativa promissora para resolver problemas com programação quadrática para estimar kt_b , com capacidade de fornecer maior precisão em comparação com outras técnicas existentes. O SMO também pode ser utilizado para a estimativa de kt_b em regiões para as quais não existem dados de radiação e tenha condições climáticas semelhantes a Botucatu-SP.

Agradecimentos

Os Autores agradecem o CNPq e a FAPESP pelo apoio financeiro.

REFERÊNCIAS

- Ångström, A. K., 1924. Solar and terrestrial radiation. Quarterly Journal of the Royal Meteorological Society, v.50, p.121.
- Beyer, H. G., Fauter, M., Schumann, K., Schenk, H., Meyer, R., 2010. Synthesis of DNI time series with sub-hourly time resolution. In: 16th SolarPACES Conference. Perpignan (France).
- Blanc, P. et al., 2014. Direct normal irradiance related definitions and applications: the circumsolar issue. Solar Energy, v.110, p.561–77.
- 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, p. 311–318.
- Chen, J-L, LI, G-S., 2014. Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. Theoretical and Applied Climatology, v.115, p.627–638.
- Chen, J-L., Li, G-S., Xiao, B-B., Wen, Z-F., LV, M-Q., Chen, C-D., Jiang, Y., Wang, X-X., Wu, S-J., 2015. Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. Energy Conversion and Management, v.89, p. 318–329.
- Codato, G., Oliveira, A. P., Soares, J., Escobedo, J. F., Gomes, E. G., Dal Pai, A., 2008. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. Theoretical and Applied Climatology, v. 93, p.57–73.
- 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.

- Fernandez-Peruchena, C., Blanco, M., Bernardos, A., 2014. Generation of series of high frequency DNI years consistent with annual and monthly long-term averages using measured DNI data. *Energy Procedia*, v.49, p. 2321–2329.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten. I. H., 2009. The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, Volume 11, Issue 1.
- 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.
- Kotti, M. C., Argiriou, A. A., Kazantzidis, A., 2014. Estimation of direct normal irradiance from measured global and correct diffuse horizontal irradiance. *Energy*, v.70, p.382-392.
- Lorena, A. C., Jacintho, L. F. O., Siqueira, M. F., Giovanni, 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.
- Mohammadi, K., Shamshirband, S., Anisi, M. H., Alam, K. A., Petkovic, D., 2015. Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion and Management*, v.91, p.433–441.
- Padovan, A., Del Col., Sabatelli, V., Marano, D., 2014. DNI estimation procedures for the assessment of solar radiation availability in concentrating systems. *Energy Procedia*, v. 57, p.1140-1149.
- PLATT, J., 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Redmond: Microsoft Research. 21 p. TechReport.
- Santos, C. M. et al., 2014. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. *Energy*, v.71, p.388-398.
- Smola, A. J., Schölkopf, B., 2004. A tutorial on support vector regression. *Statistics and Computing*, v.14, p.199-222.
- Vapnik, V. N. 1995. The nature of statistical learning theory. New York: Springer.
- Viana, T. S., Ruthera, R., Martins, F. R., Pereira, E. B., 2011. Assessing the potential of concentrating solar photovoltaic generation in Brazil with satellite-derived direct normal irradiation. *Solar Energy*, v.85, p.486–495.
- Witten, I. H., Frank, E., Hall, M. A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd ed, 630p.
- WMO - 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, p. 298.
- Yadav, A. K., Chandel, S. S., 2015. Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model. *Renewable Energy*, v.75, p.675-693.

EVALUATION OF SEQUENTIAL MINIMAL OPTIMIZATION (SMO) IN ESTIMATING BEAM SOLAR FRACTION AT NORMAL INCIDENCE TRANSMITTED (kt_b^d).

Abstract. In this work, the Sequential Minimal Optimization (SMO) algorithm is used to estimate the beam solar fraction at normal incidence transmitted on terrestrial surface (kt_b^d). The Radial Basis Function (RBF) kernel is used for regression. The statistical model (#M3) is developed and compared with the model (SMO3). The input variable used is the insolation ratio (n / N) [n is the sunshine and N the photoperiod]. 13 years of measurements were used to Botucatu - SP region. Two named database typical year (AT) and atypical year (AAT), selected of the total base of 13, are used to validate the models. In the evaluation of the models were used: Relative Mean Bias Error (rMBE), Relative Root Mean Square Error (rRMSE), percentage relative error (ϵ) and Willmott index of agreement (d). The SMO3 has better accuracy than the model (#M3). The validation with AT and AAT was satisfactory. Finally, SMO3's performance in the wet and dry season is analyzed to verify the influence of clouds, aerosols and water vapor in the dispersion of estimates and increased in the errors. The SMO estimated kt_b^d with better precision and can be used.

Key words: SVM, Insolation Ratio, WEKA.