

MÁQUINAS DE VETORES DE SUPORTE (SVM) NA ESTIMATIVA DA FRAÇÃO DIFUSA DA IRRADIAÇÃO SOLAR GLOBAL

EDSON LUIS BASSETTO¹, ALEXANDRE DAL PAI², ADRIANO DE SOUZA MARQUES³

¹ Departamento de Elétrica, Universidade Tecnológica Federal do Paraná (UTFPR) câmpus Cornélio Procópio, avenida: Alberto Carazzai 1640, Cep: 86300-000, Cornélio Procópio, Paraná, Brasil. email:bassetto@utfpr.edu.br.

² Departamento de Bioprocessos e Biotecnologia, Universidade Estadual Paulista (UNESP) câmpus Botucatu, avenida: Universitária 3780, Cep: 18610-034, Botucatu, São Paulo, Brasil. email:dal.pai@unesp.br

³ Departamento de Informática, Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) câmpus Birigui, rua: Pedro Cavallo 709, Cep: 16201-407, Birigui, São Paulo, Brasil. email:adriano.marques@ifsp.edu.br

RESUMO: A fração difusa (K_d) é a relação entre a irradiação solar difusa (H_d) e a irradiação solar global (H_g) e é um importante parâmetro para se obter medidas de H_d quando dados da H_g são conhecidos, pois permite minimizar a dependência do dia e local das informações ficando dependente somente das condições climáticas da localidade. Este trabalho tem como objetivo utilizar a técnica de Máquina de Vetores de Suporte (SVM) na estimativa da K_d na partição diária. Como entrada para aplicação da técnica são utilizadas informações de variáveis astronômicas, geográficas e meteorológicas. A técnica é comparada com modelo estatístico (ME) que utiliza como variável somente o índice de claridade (K_T). Os dados utilizados são do período de 2000 a 2006. No treinamento da técnica é utilizado um conjunto de combinações de variáveis. Para validação os dados são organizados formando uma base anual denominada de Ano Típico (AT) e de Ano Atípico (AAT) e utilizados para treinamento e elaboração do modelo. O desempenho é avaliado pelo coeficiente de correlação (r) e a raiz quadrada do erro quadrático médio (RMSE). Os resultados mostram uma melhoria no desempenho com o uso da técnica SVM quando as variáveis são inseridas de forma progressiva contribuindo com as estimativas de K_d nas condições utilizadas.

Palavras-chaves: Radiação difusa, Técnicas de aprendizado de máquinas, Método de estimativa.

SUPPORT VECTOR MACHINES (SVM) IN GLOBAL SOLAR IRRADIATION DIFFUSED FRACTION ESTIMATE

ABSTRACT: Diffuse fraction (K_d) is the relationship between diffuse solar irradiation (H_d) and global solar irradiation (H_g) and is an important parameter for obtaining measurements of H_d when data H_g are known because it allows to minimize the day dependence and information place, being dependent only on the locality climatic conditions. This work aims to use the Support Vector Machine (SVM) technique to estimate K_d in the daily partition. As input for the application of the technique was use information of astronomical, geographical and meteorological variables. The technique is compared with a statistical model (ME) that uses only the lightness index (K_T) as a variable. The data used are from 2000 to 2006. The technique training uses a set of combinations of variables. For validation, the data are reorganized into an annual basis called Typical Year (TY) and Atypical Year (AAY) and independent of data for training and model development. Performance is assessed by the correlation coefficient (r) and the square root of the mean square error (RMSE). The results show an improvement in performance using the technique SVM when the variables are entered progressively contributing to the estimates of K_d with the conditions used.

Keywords: Diffuse radiation, Machine learning techniques, Estimation method.

1 INTRODUÇÃO

Considerada como uma energia limpa, sustentável e abundante, a irradiação solar tem importância nas áreas científica e tecnológica. Contudo, para um bom aproveitamento dessa energia é relevante quantificar ou estimar esse potencial por meio de medidas (séries temporais ou espaciais) ou através de modelos que possam estimá-la e com isso verificar e avaliar sua disponibilidade ao longo do tempo (Bakirci, 2015; Teke et al., 2015).

Dos dados obtidos das componentes da radiação solar, as medidas da irradiação solar global ($H_g - \text{MJ/m}^2$) é a componente monitorada com maior frequência nas estações solarimétricas, diferentemente das medidas da irradiação solar direta ($H_b - \text{MJ/m}^2$) e irradiação solar difusa na horizontal ($H_d - \text{MJ/m}^2$) que, para suas obtenções, demandam de equipamentos custosos e de manutenções periódicas.

Geralmente H_b é obtida pela diferença entre H_g e H_d por meio da relação $H_b = H_g - H_d$, pois a obtenção das medidas de H_b apresenta custos elevados de equipamentos, manutenção e operação (Santos et al., 2016a; Dal Pai et al., 2016), diferentemente das medidas de H_g , que são comumente monitoradas, e de H_d que podem ser obtidas através do método do anel ou disco de sombreamento com um custo menor (Melo e Escobedo, 1994; Marques Filho et al., 2016; Drechmer e Ricieri 2006; Codato et al., 2008; Furlan, et al., 2012; Soares et al., 2004).

Medidas de H_d ainda são escassas e com poucas informações disponíveis, mesmo com sua importância. (Silva et al., 2008; Oliveira et al., 2002a; Santos et al., 2016). Diferentemente da obtenção de dados medidos ou calculados dessa irradiação H_d , os modelos existentes, como o de estimativa, precisam de informações na superfície terrestre, como os modelos estatísticos e as técnicas de aprendizado de máquinas (Jamil e Akhatar, 2017; Khorasanizadeh et al., 2016; Ruiz-Arias et al., 2010).

Geralmente os modelos para estimar a H_d são modelos estatísticos e, para minimizar os efeitos de variações temporais, são propostos nestes modelos o uso de frações que permitem minimizar alguns efeitos astronômicos e geográficos, como proposto por Liu e Jordan

(1960), que correlacionaram a fração difusa diária ($K_d = H_d/H_g$) com o índice de claridade ($K_T = H_g/H_0$, onde $H_0 - \text{MJ/m}^2$ é a irradiação no topo da atmosfera). Outros modelos foram desenvolvidos na literatura em diferentes partições de tempo (Collares-Pereira e Rabl, 1979; Jacovides et al., 2006; Paleuscu e Blaga, 2016), como também modelos com mais de uma variável (Reindl et al., 1990; Furlan et al., 2012).

Além disso, nos últimos anos surgiram modelos de estimativas alternativos chamados Técnicas de Aprendizado de Máquinas (TAM), como as Redes Neurais Artificiais (RNAs), as Máquinas de Vetores de Suporte (SVMs) E DE técnicas híbridas que utilizam das Redes Neurais e os Sistemas de Inferência Fuzzy. Essas técnicas permitem utilizar variáveis de grandezas físicas diferentes e prever medidas a partir de informações conhecidas e desejadas (Khatib et al, 2012; Paulescu et al, 2016).

Com as considerações descritas e da importância de conhecer medidas de H_d na superfície horizontal, pretende-se utilizar a técnica SVM na estimativa da fração difusa K_d na partição diária, inserindo progressivamente variáveis climatológicas na modelagem como base em dados obtidos na cidade de Botucatu (SP) e comparar as estimativas de H_d com os resultados apresentados do modelo estatístico de referência, elaborado com a mesma base de dados e com apenas uma variável independente.

2 MATERIAL E MÉTODOS

2.1 Localização e clima

Os dados para realização deste trabalho foram obtidos na Estação Meteorológica Lageado, que está localizada na Faculdade de Ciências Agrônomicas do campus da UNESP – Universidade Estadual Paulista - em Botucatu (SP), Brasil. O município tem temperatura do ar média anual de 20,46°C e umidade relativa do ar variando de 61,20% (agosto) a 76,50% (janeiro). A evolução da nebulosidade, que conforme a WMO (2008) em seu guia de instrumentos meteorológicos e métodos de observação é o limiar para o brilho do sol na superfície e segue a variação da temperatura sendo mais elevada nos meses do verão (61%) e

menor nos meses de inverno (28%). A precipitação é constituída de dois intervalos distintos com um período chuvoso com maior precipitação no verão, com média total de 304 mm e o período seco com menor valor no inverno, com média de 38 mm.

2.2 Obtenção da base de dados

Os dados foram obtidos simultaneamente, no período de 2000 a 2006, e consiste de medidas instantâneas (W/m^2) da irradiância solar global ($I_g - W/m^2$), da irradiância solar difusa na horizontal ($I_d - W/m^2$), da irradiância solar direta normal ($I_{bn} - W/m^2$), da irradiância solar de ondas longas ($I_{wa} - W/m^2$), da temperatura do ar ($tar - ^\circ$) e da umidade relativa ($umid - \%$). Para as medidas de I_g e I_d foram utilizados dois piranômetro Eppley-PSP sendo que para medidas da irradiação difusa foi aplicado o método com anel de sombreamento (Melo e Escobedo, 1994). Para medidas de I_{bn} foi usado um pireliômetro Eppley-NIP acoplado a um rastreador solar do tipo ST-3 e calculada para irradiância direta projetada na horizontal (I_b) através da Equação 1:

$$I_b = I_{bn} * \cos(\theta_z) \quad (1)$$

Onde θ_z é o ângulo zenital (graus). Para medidas de I_{wa} , um pirômetro, modelo CG1 da Kipp e Zonen e para ($umid$) um termógrafo e um higrógrafo da Filotécnica Milano, instalados dentro de um abrigo meteorológico. Neste estudo, os valores da irradiância difusa I_d foram corrigidos para compensar a parcela de I_d obstruída pelo próprio anel, a qual depende de parâmetros geométricos, geográficos e astronômicos (Oliveira et. al, 2002b). Os equipamentos são aferidos pelo método comparativo todos os anos, através de instrumentos específicos.

Na aquisição e armazenamento dos dados foi utilizado um sistema de aquisição automático digital (*datalogger*) Campbell Cr23X da Campbell Scientific, onde os valores foram armazenados e transmitidos para um computador. Após obtenção das medidas instantâneas (unidade - W/m^2), os dados foram integrados na partição diária (unidade - MJ/m^2)

(Chaves e Escobedo, 2000). As medidas passaram por um controle de qualidade e foram retirados aproximadamente 8,5% do total de dados medidos do período.

2.3 Modelo estatístico - ME

O modelo estatístico foi elaborado com valores de K_d e de K_T , calculados a partir de dados experimentais e foram correlacionados (Liu e Jordan, 1960), sendo a equação de estimativa ajustada por regressão polinomial de 4ª ordem conforme modelo apresentado na Equação 2.

$$K_d = \beta_0 + \beta_1 K_T + \beta_2 K_T^2 + \dots + \beta_n K_T^n \quad (2)$$

2.4 Máquina de vetores de suporte - SVM

A técnica SVM foi desenvolvida por Vapnik (1995) para resolver problemas de classificação e aplicável para regressão e estimativas (geralmente chamada de Regressão por Vetores de Suporte). É diferenciada dos modelos estatísticos por utilizar o princípio da minimização do erro estrutural e, com isso, diminuir o erro empírico e o intervalo de confiança com grande capacidade de generalização (Antonanzas-Torres et al., 2015; Santos et al., 2016b).

O conceito básico da técnica é realizar o treinamento para aprendizagem da máquina através de algoritmos de otimização matemática envolvendo poucos parâmetros de ajuste para seu treinamento e que depende da sua utilização e, conseqüentemente, do algoritmo de treinamento a ser utilizado (supervisionado). A proposta da SVM é criar uma margem e maximizá-la para realizar o mapeamento dos dados de entrada e do treinamento em um espaço de dimensões maiores via funções de *kernel* que ajudam o algoritmo a ficar mais eficiente, construindo um hiperplano com uma margem de separação máxima entre os dados. Matematicamente, a SVM elabora um hiperplano de separação ideal considerando um conjunto de dados (x_i, y_i) chamados de vetores de entrada (x_i) e seus respectivos vetores de saída (y_i) normalizados (Belaid e Mellit, 2016) em um espaço através de uma função não linear

$\varphi(x)$ e na sequência encontrar uma função de regressão conforme Equação 3:

$$f(x) = \omega \cdot \varphi(x) + b \quad (3)$$

Onde ω e b são os parâmetros da função de regressão chamados respectivamente de vetor de ponderação (pesos) e valor de tendência (bias) e $\varphi(x)$ como função de mapeamento não linear. Com isto, a programação irá encontrar um peso (ω) e um limiar (b) que tem a função de otimizar as margens de dados utilizados para estimativa e aproximar a saída (y) com um tolerância de erro (ϵ) a fim de assegurar uma boa classificação ou regressão com o menor erro da $\varphi(x)$. A (SVM) calcula o hiperplano de separação ideal por meio do seguinte problema de otimização dado pela Equação 4:

$$R(\xi, \omega, b) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \quad (4)$$

Sendo $\xi_i \geq 0, i = 1, \dots, N$ as variáveis de folga utilizadas para reduzir o erro de classificação do conjunto de treinamento; C o parâmetro de custo que determina o grau do erro empírico no problema de otimização escolhido pelo usuário. As variáveis de folga também possibilitam à SVM tratar de problemas não lineares. Sem o emprego das variáveis de folga, a SVM é chamada de “SVM de margens rígidas” e, com as variáveis de folga, passa a se chamar de “SVM de margens flexíveis”. Para resolver problemas de otimização ou de programação quadrática, os Multiplicadores de Lagrange são incorporados às equações de condição e pode ser escrita de forma dual e introduzindo a função de *Kernel* permitindo que a SVM tenha a capacidade de modelar hiperplanos de separação complicados, ou seja, forme limites não lineares. Essa característica de maximização de espaço através das funções de

Kernel é que pode fazer com que a técnica SVM seja viável para abordar estudos com irradiação solar como H_g (Mohammadi et al., 2015b; Antonanzas-Torres et al., 2015; Belaid e Mellit, 2016; Chen et al., 2015) e a irradiação difusa (Shamshirband et al., 2016). Apesar da técnica SVM ser pouco utilizada em estimativas de irradiação solar, no hemisfério sul vem sendo introduzida na estimativa de H_b (Santos et al., 2016b), na H_g (Silva et al., 2017) e irradiação solar Ultra Violeta (Teramoto et. al., 2016).

2.5 Treinamento da técnica SVM

Na implementação da técnica SVM, os dados foram submetidos para execução no *software* Waiko Environment for Knowledge Analysis - WEKA (Hall et al., 2009), ferramenta KDD – Knowledge Discovery in Databases, que agrega uma série de algoritmos de preparação de dados, aprendizagem e de validação de resultados de forma interativa. No treinamento e teste foi utilizado o algoritmo de regressão SVM/SMO – Otimização Mínima Sequencial com função de transformação de base radial (RBF) de *Kernel* com método de validação cruzada e submetido à validação de acordo com o conjunto de combinações e dados para parametrização, apresentados na Tabela 4.

Os parâmetros da configuração básica do método SVM/SMO, como o parâmetro de margem (C), que determina o ponto de equilíbrio entre a maximização da margem e a minimização do erro e que controla a influência de cada vetor; o parâmetro (γ), que controla a flexibilidade da função de *Kernel* e verifica seu ajustamento (*overfitting e underfitting*); o parâmetro (ϵ) erros de tolerância são definidos testando-se todos os valores de acordo com a Tabela 1 e determinados empiricamente.

Tabela 1. Valores para parametrização da SVM

C	1	10	50	100	200	300
γ	0,1	0,2	0,3	0,4	0,5	0,6
ϵ	0,05	0,05	0,05	0,05	0,05	0,05

2.6 Elaboração e validação dos modelos

Para elaboração do modelo ME e treinamento da técnica SVM, os dados diários foram separados de toda a série entre 2000 a 2006, e a partir da composição de duas bases anuais utilizadas para validação dos modelos, denominadas de AT e AAT.

Esse arranjo foi estruturado a partir de uma análise estatística da média dos valores da H_d de cada mês do conjunto de anos. Após este

estudo, foram comparados, para cada mês entre os anos, os valores mais próximos da média dos valores da H_d como sendo meses típicos e os valores mais afastados da média como atípicos, formando assim uma base de AT e outra de AAT de medidas diárias. Na Tabela 2 são apresentados os quantitativos de dados para elaboração do modelo ME (a), do treinamento da técnica SVM (b) e para validação de ambos os modelos (c).

Tabela 2. Dados para elaboração e validação modelos

Total de dados: 2348 = 100%		
a) Elaboração do modelo (ME)	1.703 – 72,53%	
b) Treinamento da técnica (SVM)	Base AT	Base AAT
	2.024 – 86,20%	2.027 – 86,33%
c) Validação modelo (ME) e técnica (SVM)	Base AT	Base AAT
	324 – 13,80%	321 – 13,67%

Na Tabela 3 são apresentados os meses de cada ano mais próximo da média como sendo AT e os mais afastados como AAT.

Como exemplo, o mês de janeiro que ficou mais próximo da média foi o do ano de

2003 e julho, o do ano de 2005 e, para os mesmos meses, os anos que ficaram mais afastados da média foram 2002 e 2004.

Tabela 3. Base Ano Típico (AT) e Ano Atípico (AAT)

Mês/Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Dez
AT	03	00	01	01	00	01	05	01	06	06	05
AAT	02	03	00	04	04	02	04	00	05	05	01

Na geração do modelo ME, somente os dados das variáveis K_T e K_d são utilizados. Para treinamento da técnica SVM foi definida uma sequência de combinações ou arranjos com

todas as variáveis climatológicas medidas e/ou calculadas disponíveis como entrada para técnica (Tabela 4).

Tabela 4. Combinações e variáveis de entrada

Combinações	Variáveis de:										
	Entrada										Saída
Comb 1	K_T										
Comb 2	K_T	H_0									
Comb 3	K_T	H_0	H_g								
Comb 4	K_T	H_0	H_g	H_{wa}							K_d
Comb 5	K_T	H_0	H_g	H_{wa}	decsol						
Comb 6	K_T	H_0	H_g	H_{wa}	decsol	nebul					
Comb 7	K_T	H_0	H_g	H_{wa}	decsol	nebul	umid				
Comb 8	K_T	H_0	H_g	H_{wa}	decsol	nebul	umid	tar			

Medido e calculado

Medido

Calculado

Essas variáveis consistem em valores medidos de H_g , H_{wa} , $umid$ e tar e calculados como H_0 , K_T , K_d declinação solar ($decsol - ^\circ$) e nebulosidade ($nebul - \%$), que foi calculada em função da proporcionalidade das medidas da irradiância direta I_b dos valores abaixo de 120 W/m^2 . Conforme a WMO (2008), em seu guia de instrumentos meteorológicos e métodos de observação, este valor de I_b foi adotado com limiar para o brilho do sol, ou seja, para valores inferiores a esses, considera-se sem insolação. Outros arranjos foram analisados, porém em função das condições de seleção das variáveis não foram utilizados.

Neste estudo, optou-se em utilizar variáveis considerando sua disponibilidade, importância, período de medidas e características físicas em relação à variável a ser estimada. Não foi realizada uma análise estatística para reduzir ou eliminar sobreposições de valores, nem para verificar as mais representativas. Definidas as combinações das variáveis de entrada apresentam-se na sequência a metodologia para treinamento da técnica.

2.7 Indicadores estatísticos

Na análise dos resultados foram utilizados como indicadores estatísticos o coeficiente de determinação (R^2) para ajuste do modelo ME, a raiz quadrada do desvio quadrático médio (RMSE - *Root Mean Square Error*), o erro quadrático médio (MSE - *Mean Square Error*) e o coeficiente de correlação (r) para indicadores de treinamento e validação do modelo ME e da técnica SVM. O coeficiente de determinação permite verificar o quanto a variação da variável dependente é explicada

pela variação da variável independente. A raiz quadrada do desvio quadrático médio permite verificar o valor do erro produzido pelo modelo, sendo valores menores indicando um desempenho satisfatório para os modelos gerados, testados ou treinados e validados, ou seja, mostra o desempenho do modelo quanto ao seu espalhamento. O erro quadrático médio possibilita verificar o erro quadrático médio entre os valores estimados e medidos, ou seja, os melhores resultados são aqueles com desvios menores mais próximos dos valores desejados, e o r que permite verificar o quanto o valor medido se aproxima do valor estimado, ou seja, a correlação existente entre as variáveis (Stone, 1993; Escobedo et al. 2011).

3 RESULTADOS E DISCUSSÃO

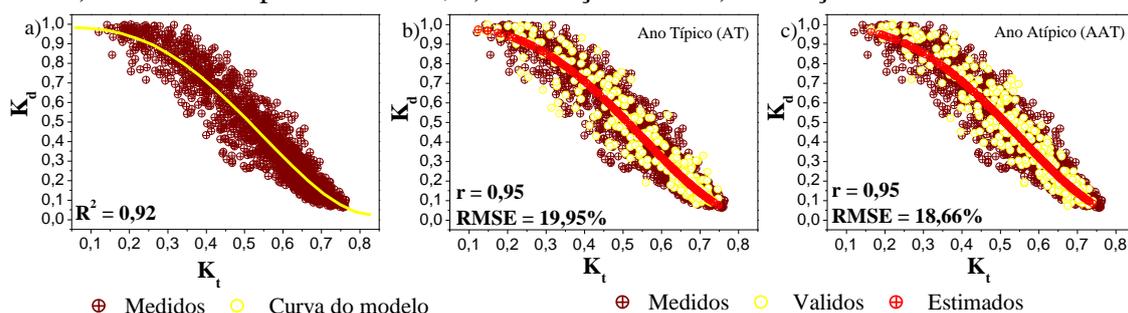
3.1 Modelo estatístico

O modelo estatístico da correlação de K_d em função de K_T (Figura 1 (a)) apresenta uma curva similar às curvas encontradas na literatura, assim como as equações de estimativa (Equação 5) ajustada por regressão polinomial de 4ª ordem e com coeficiente de determinação na ordem $R^2 = 0,92$, o que demonstra que K_d está correlacionado com K_T estatisticamente. Outras equações de ordens polinomiais foram ajustadas, mas apresentaram coeficiente de determinação mais baixo.

$$K_d = 0,98 + 0,089K_T + 0,46K_T^2 + 5,88K_T^4 \quad (5)$$

São apresentadas na Figura 1 (b) e (c) as curvas de dispersão geradas na validação do modelo ME para AT e AAT, com os valores dos dados medidos (marrom), dos dados de validação (amarelo) e da estimativa (vermelho).

Figura 1. a) Curva de dispersão do ME; b) Validação AT e c) validação AAT



O modelo ME na base AT apresentou uma correlação satisfatória ($r = 0,95$) e um erro quadrático aceitável (RMSE = 19,95%), assim como na base de validação AAT, que também apresentou uma correlação satisfatória ($r = 0,95$) e um erro quadrático bom (RMSE = 18,66%), mostrando que o modelo pode ser utilizado na estimativa da fração difusa na partição diária, utilizando somente K_T como variável.

Na Tabela 5 são apresentados os resultados encontrados do (r) e de RMSE da literatura pesquisada com o melhor resultado do modelo estatístico ME para partição diária.

Os resultados mostram valores similares ou até superiores em indicadores de correlação entre os modelos, mostrando que o modelo ME pode ser utilizado na estimativa de K_d , tendo em vista que a elaboração dos modelos possuem as mesmas características construtivas.

Tabela 5. Índices estatísticos - outras localidades e ME para Botucatu

Autores	Localidade	r	RMSE %
Khorasanizadeh et al. 2016	Irã	0,88	
Bortolini et al. 2013	Madri	0,96	
Shamshirband et al. 2016	Irã	0,82	
Oliveira et al. 2002a	São Paulo	0,90	
ME AT – Proposto 2017	Botucatu/São Paulo	0,95	19,95
ME AAT – Proposto 2017	Botucatu/São Paulo	0,95	18,66

3.2 Técnica SVM

Os resultados, após treinamento e validação com as combinações (Tabela 4) da técnica SVM, são apresentados na Tabela 6 com os melhores indicadores experimentados na parametrização com as duas bases de validação.

Pode-se observar pelos indicadores de K_d que o índice de correlação r entre os valores medidos e estimados nas duas bases de validação com a SVM foi de 0,95 a 0,96, mostrando uma boa correlação tanto para AT quanto para AAT. O treinamento da técnica

SVM com o algoritmo SMO, depende do ajustamento das margens dos vetores de suporte que regula a quantidade de vetores e conseqüentemente do número de amostras e variáveis que participam do treinamento. São apresentados, na Tabela 6, a configuração e parametrização das melhores combinações, o termo de maximização das margens entre os vetores de suporte, o erro de tolerância e de desvio e os indicadores estatísticos dos valores estimados em função dos valores medidos na validação.

Tabela 6. Indicadores e configurações da técnica SVM

	Para Ano típico (AT)							
	Coeficiente de parametrização			Indicadores estatísticos				
	C	γ	ϵ	Treinamento		Validação		
MSE				RMSE (%)	MSE	RMSE (%)	r	
Comb 1	100	0,2	0,05	0,29	30,32	0,0067	20,04	0,95
Comb 2	200	0,2	0,05	0,28	29,78	0,0065	19,69	0,95
Comb 3	300	0,2	0,05	0,28	29,80	0,0065	19,70	0,95
Comb 4	100	0,2	0,05	0,28	29,55	0,0064	19,54	0,96
Comb 5	100	0,2	0,05	0,28	29,72	0,0064	19,66	0,95
Comb 6	100	0,2	0,05	0,28	29,90	0,0061	19,12	0,96
Comb 7	100	0,2	0,05	0,28	28,86	0,0061	19,08	0,96
Comb 8	300	0,2	0,05	0,27	27,69	0,0056	18,32	0,96

	Para Ano atípico (AAT)							
	Coeficiente de parametrização			Indicadores estatísticos				
	C	γ	ϵ	Treinamento		Validação		
MSE				RMSE (%)	MSE	RMSE (%)	r	
Comb 1	200	0,2	0,05	0,29	30,55	0,0069	18,53	0,95
Comb 2	300	0,2	0,05	0,28	30,15	0,0067	18,30	0,95
Comb 3	300	0,2	0,05	0,29	30,20	0,0067	18,33	0,95
Comb 4	300	0,2	0,05	0,29	29,99	0,0066	18,20	0,95
Comb 5	300	0,2	0,05	0,28	29,81	0,0067	18,34	0,95
Comb 6	300	0,2	0,05	0,28	28,97	0,0060	17,29	0,96
Comb 7	300	0,2	0,05	0,28	28,93	0,0059	17,18	0,96
Comb 8	50	0,1	0,05	0,27	27,38	0,0055	16,62	0,96

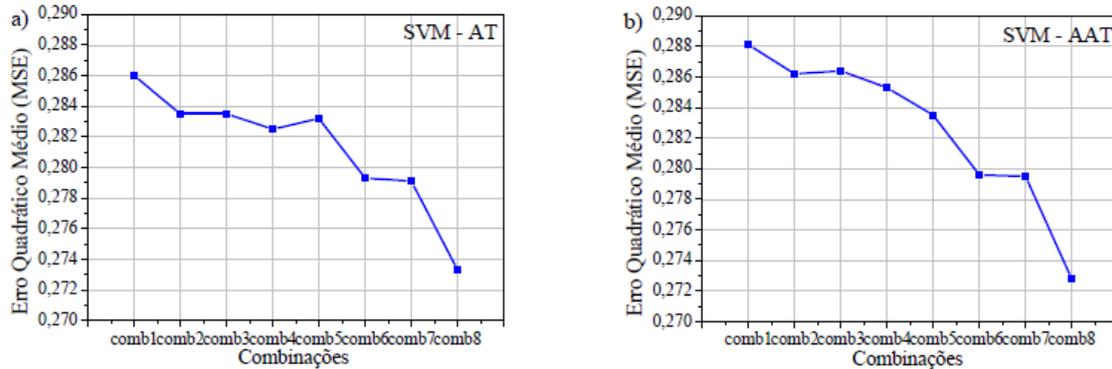
Os indicadores de desempenho RMSE nas duas bases de validação da técnica SVM ficou entre os valores 16,62% e 20,04% em todas as combinações. É possível verificar uma redução dos indicadores quando outras variáveis são inseridas no treinamento da técnica (SVM).

Os melhores resultados na validação com a base AT e AAT foram quando oito variáveis de entrada participaram do treinamento. Através da Figura 2 (a) e (b) é possível observar que, conforme as variáveis de entrada são inseridas no treinamento, o MSE diminui generalizando o problema para este caso e parametrização específico. A inserção de

outras variáveis no treinamento das redes reduz o MSE proporcionando um melhor desempenho no treinamento.

Outras parametrizações de ajustamento foram testadas, mas não resultaram em melhores indicadores de desempenho como o MSE e RMSE. Analisando os indicadores da Tabela 6 em relação aos indicadores de desempenho do modelo ME, a técnica SVM, apresenta um desempenho melhor com uma redução de 10% em comparação ao modelo ME. A inserção de outras variáveis no treinamento contribuí para um melhor desempenho com uma diminuição do MSE de treinamento.

Figura 2. Curva erro quadrático médio (MSE) x combinações - validação para base a) AT e b) AAT.

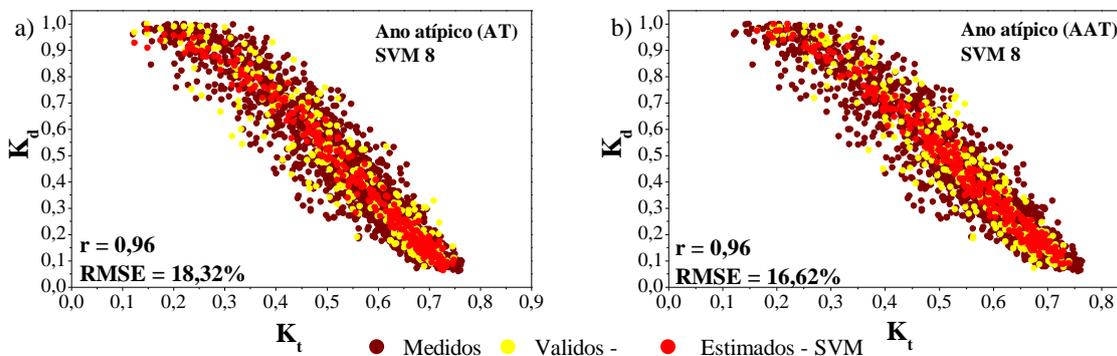


Para um melhor entendimento de aproximação entre os valores medidos e estimados, a Figura 3 (a) e (b) mostra o espalhamento dos dados da melhor combinação (oito variáveis) para estimativa de K_d nas duas bases AT e AAT.

Deve-se ressaltar que os resultados dos indicadores estatísticos encontrados neste trabalho são referentes às combinações apresentadas na Tabela 4. A exclusão ou inclusão de variáveis e das quantidades de dados

utilizados podem influenciar o ajustamento dos parâmetros da técnica SVM, como no parâmetro C (margem) que determina o ponto de equilíbrio entre a maximização da margem dos vetores de suporte e a minimização do erro que é influenciado sempre que os vetores (dados) se alteram. Outras combinações e ajustamentos podem ser testadas a fim de verificar o comportamento e desempenho da técnica em seus treinamentos.

Figura 3. Curva dispersão K_T x K_d : a) AT e b) AAT



Os resultados apresentados mostram que a técnica SVM é indicada na estimativa de K_d com o uso de variáveis climatológicas e meteorológicas com indicadores superiores ao modelo ME, sendo uma alternativa para a estimativa da fração difusa da irradiação solar global.

4 CONCLUSÕES

Diante do estudo realizado neste trabalho, conclui-se que o modelo ME com $R^2 = 0,92$ pode ser utilizada para estimar a fração difusa da irradiação global com valores de

validação de $RMSE_{AAT} = 19,95\%$ e $RMSE_{AAT} = 18,66\%$ similares aos encontrados na literatura pesquisada e que a técnica SVM apresenta valores mais baixos do MSE nos treinamentos em todas as combinações com inserções de variáveis.

A validação da técnica SVM mostra que os resultados (valores de RMSE) são similares e superiores ao modelo ME quando está na mesma condição de variáveis de entrada, que neste caso, somente K_T . Com a inserção progressiva das variáveis climatológicas e com as parametrizações testadas, os resultados proporcionaram um desempenho melhor, além

do RMSE da técnica SVM ser superior ao modelo ME na maioria dos arranjos.

Sendo assim, pode-se concluir que a técnica SVM pode ser utilizada como mais uma alternativa dos modelos de estimativa da fração difusa da irradiação solar global.

5 AGRADECIMENTOS

À Universidade Tecnológica Federal do Paraná - UTFPR e a Fundação Araucária/CAPES pelo apoio total e parcial para realização desse trabalho.

6 REFERÊNCIAS

- BAKIRCI, K. Models for the estimation of diffuse solar radiation for typical cities in Turkey. **Energy**. vol. 82, p. 827 – 838, 2015.
- BELAID, S.; MELLIT, A. Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. **Energy Conversion and Management**, vol. 118, p. 105–118, 2016.
- BORTOLINI, M. et al. Multi-location model for the estimation of the horizontal daily diffuse fraction of solar radiation in Europe. **Energy Conversion and Management**, vol. 67, p. 208–216, 2013.
- CHAVES, M.; ESCOBEDO, J. F. A software to process daily solar radiation data, **Renewable Energy**, vol.19(1), p. 339 – 344, 2000.
- CHEN, J. L. et al. Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. **Energy Conversion and Management**, vol. 89, p. 318–329, 2015.
- CODATO, G. et al. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. **Theoretical and Applied Climatology**, vol. 93, p. 57–73, 2008.
- COLLARES-PEREIRA, M.; RABL, A. The average distribution on solar radiation – correlations between diffuse and hemispherical and between daily and hourly insolation values. **Solar Energy**, vol. 22, p. 155 - 164, 1979.
- DAL PAI, A. et al. MEO shadowring method for measuring diffuse solar irradiance: corrections based on sky cover. **Renewable Energy**, vol. 99, p.754 - 76, 2016.
- DRECHMER, P. A. O.; RICIERY, R. P. Irradiação global, direta e difusa, para a região d Irradiação global, direta e difusa, para a região de para a região de Cascavel, Estado e Cascavel, Estado do Paraná do Paraná. **Acta Scientiarum Technology**, vol. 28, n. 1 p. 73 – 77, 2006.
- ESCOBEDO, J. F. et al. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. **Renewable Energy**, v.36, p.169-178, 2011.
- FURLAN, C. et al. The role of clouds in improving the regression model for hourly values of diffuse solar radiation. **Applied Energy**, vol. 92, p. 240 – 254, 2012.
- HALL, M. et al. The WEKA Data Mining Software: An Update; SIGKDD Explorations, v.11, Issue 1, 2009.
- JACOVIDES, C. P. et al. Comparative study of various correlations in estimating hourly diffuse fraction of global solar radiation. **Renewable Energy**, vol. 31, p. 2492 – 2504, 2006.

JAMIL, B.; AKHTAR, N. Estimation of diffuse solar radiation in humid-subtropical climatic region of India: Comparison of diffuse fraction and diffusion coefficient models. **Energy**, vol. 131, p. 149 – 164, 2017.

KHATIB, T. et al. Assessment of Artificial Neural Networks for Hourly Solar Radiation Prediction. **International Journal of Photoenergy**, vol. 2012, p. 1 - 8, 2012.

KHORASANIZADEH, H.; MOHAMMADI, K.; GOUDARZI, N. Prediction of horizontal diffuse solar radiation using clearness index based empirical models: A case study. **International Journal of Hydrogen Energy**, vol. 41, p. 21888 – 21898, 2016.

LIU, B. H. Y.; JORDAN, R. C. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. **Solar Energy**, vol. 4, p. 1, 1960.

MARQUES FILHO, E. P. et al. Global, diffuse and direct solar radiation at the surface in the city of Rio de Janeiro: Observational characterization and empirical modeling. **Renewable Energy**, vol. 91, p. 64 – 74, 2016.

MELO, J. M. D.; ESCOBEDO, J. F. Medida da radiação solar difusa. In: Energias Limpas em progresso. VII Congresso Ibérico de Energia Solar, Vigo, Espanha. Anais **International Solar Energy Society**, vol. 1, 1994.

MOHAMMADI, K. et al. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. **Energy Conversion and Management**, vol. 92, p. 162–171, 2015b.

OLIVEIRA, A. P. et al. Correlation models of diffuse solar-radiation applied to the city of São Paulo, Brazil. **Applied Energy**, vol. 71, p. 59 – 73, 2002a.

OLIVEIRA, A. P., ESCOBEDO, J. F., MACHADO, A. J. A new shadow-ring device for measuring diffuse solar radiation at surface. **Journal Atmos. Ocean. Technol.** Vol.19, p. 698 – 708, 2002b.

PAULESCU, E.; BLAGA, R. Regression models for hourly diffuse solar radiation. **Solar Energy**, vol. 125, p. 111 – 124, 2016.

REINDL, D. T.; BECKMAN W. A.; DUFFIE, J. A. Diffuse fraction correlations. **Solar Energy**, v.45, p.1 – 7, 1990.

RUIZ-ARIAS, J. A. et al. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. **Energy Conversion and Management**, vol. 51, p. 881 – 893, 2010.

SANTOS C. M.; ESCOBEDO J. F. Temporal variability of atmospheric turbidity and DNI attenuation in the sugarcane region, Botucatu/SP. **Atmospheric Research**, vol. 181, p. 312- 321, 2016a.

SANTOS, C. M. et al. Assessment of ANN and SVM models for estimating normal direct irradiation (H_b). **Energy Conversion and Management**, vol. 126, p. 826–836, 2016b.

SHAMSHIRBAND, S. et al. Estimating the diffuse solar radiation using a coupled support vector machine–wavelet transform model. **Renewable and Sustainable Energy Reviews**, vol. 56, p.428–435, 2016.

SILVA, F. R.; OLIVEIRA, H. S. M.; MARINHO, G. S. Análise das componentes global e difusa da radiação solar em Natal-RN entre 2007 e 2008. **II Congresso Brasileiro de Energia Solar e III Conferência Regional Latino-Americana da ISES** - Florianópolis, 18 a 21 de novembro de 2008.

SILVA, M. B. P. et al. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar–Terrestrial Physics**, vol. 160, p. 11–23, 2017.

SOARES, J. et al. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique. **Applied Energy**, vol. 79, p. 201-214, 2004.

STONE, R. J. Improved statistical procedure for the evaluation of solar radiation estimation models. **Solar Energy**, vol. 51, n^o4, p. 289 – 291, 1993.

TEKE, A.; YILDIRIM, H. B.; CELIK, O. Evaluation and performance comparison of diferente models for the estimation of solar radiation. **Renewable and Sustainable Energy Reviews**, vol. 50, p. 1097 – 1107, 2015.

TERAMOTO, E. T. et al. Uso de modelos estatísticos e de máquina de vetores de suporte na estimativa da radiação solar global diária em iguape, vale do ribeira, são paulo. **Avances en Energías Renovables y Medio Ambiente**, vol. 20, p. 11.17-11.24, 2016.

VAPNIK, C. N. Support-vector networks. **Mach Learn**, vol.20, p. 3 – 273, 1995.

WMO – World Meteorological Organization. Guide to meteorological Instruments and Methods of Observation. **WMO-nº8, Seventh Edition**, p.1 - 681, Geneva, Switzerland, 2008.