

Artigo

Comparação de Métodos de Estimativa da Radiação Solar Ultravioleta Horária: Modelos Empíricos, Redes Neurais Artificiais e Máquina de Vetores de Suporte

Érico Tadao Teramoto¹ , Cícero Manoel dos Santos², João Francisco Escobedo³, Alexandre Dal Pai³, Silvia Helena Modenese Gorla da Silva¹

¹*Campus Experimental da Unesp em Registro, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Registro, SP, Brasil.*

²*Campus Universitário de Altamira, Universidade Federal do Pará, Altamira, PA, Brasil.*

³*Faculdade de Ciências Agrárias de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Registro, SP, Brasil.*

Recebido em: 4 de Janeiro de 2019 - Aceito em: 18 de Março de 2019

Resumo

No presente trabalho, a comparação de três dos principais métodos de estimativa da radiação solar foi realizada: modelos empíricos, Redes Neurais Artificiais (RNA) e Máquina de Vetores de Suporte (SVM). Quatro modelos empíricos considerados clássicos foram calibrados e validados para a estimativa da radiação solar UV horária em Botucatu, Estado de São Paulo, Brasil. Mantendo os modelos empíricos como referências em relação a precisão e variáveis de entrada utilizadas, os desempenhos da RNA e da SVM foram avaliados. Por meio dos parâmetros estatísticos Erro Médio Padrão (MBE) e Erro Médio Absoluto (MAE) foi confirmada a superioridade da SVM em relação à RNA e aos modelos empíricos. Comparada à RNA, a SVM foi capaz de gerar melhores resultados utilizando um número menor de variáveis de entrada. Dentre todos os métodos de estimativa testados, a SVM utilizando o conjunto de variáveis de entrada $\{UV_0, K_T\}$ é considerada a melhor alternativa devido ao pequeno número de variáveis de entrada utilizadas e pela relativa acurácia.

Palavras-chave: radiação solar, aprendizado de máquinas, seleção de variáveis de entrada.

Comparing Different Methods for Estimating Hourly Solar Ultraviolet Radiation: Empirical Models, Artificial Neural Network and Support Vector Machine

Abstract

In the present paper, the comparison of three of the main estimation methods of solar radiation was performed: empirical models, Artificial Neural Network (ANN) and Support Vector Machine (SVM). Four classical empirical models were calibrated and validated in order to estimate hourly solar UV data in Botucatu, São Paulo State, Brazil. Taken the empirical models as reference of accuracy and set for input variables, the performance of ANN and SVM were assessed. Through the statistical parameters Mean Bias Error (MBE) and Mean Absolute Error (MAE) was confirmed the superiority of the SVM over the ANN and empirical models. The SVM is capable to generate better results than ANN using a less number of input variables. Among all estimation methods, SVM using the set of input variables $\{UV_0, K_T\}$ is considered the best alternative due to the smaller number of input variables and relative precision.

Keywords: solar radiation, machine learning, input variables selection.

1. Introdução

A radiação ultravioleta (UV , 280-380 nm) compõe de 4 a 6% da radiação solar incidente sobre a superfície terrestre (R_G , 280-3000 nm) (Jacovides *et al.*, 2006; Escobedo *et al.*, 2011). Porém, ela possui notoriedade devido aos seus efeitos nocivos sobre a biosfera (Huang *et al.*, 2011; Escobedo *et al.*, 2011; Wang *et al.*, 2015). Dentre os danos observados em organismos vivos devido à exposição excessiva à UV solar, podemos citar: redução no desenvolvimento em crustáceos, peixes, algas e anfíbios; redução na área foliar e na produção de frutos em plantas terrestres; eritema, câncer de pele e a catarata em seres humanos (Sucré *et al.*, 2012; Liu *et al.*, 2013; Hideg *et al.*, 2013; Kataria *et al.*, 2014; Lipinski *et al.*, 2016). Além dos efeitos nocivos em organismos vivos, a radiação UV também causa fotodegradação de fibras naturais, polímeros sintéticos e tintas. Portanto, registros confiáveis da radiação solar possui importância em pesquisas nas áreas de agricultura, aquicultura, engenharias, ecologia e ciências médicas. Contudo, o monitoramento dela ainda permanece ausente em muitas estações meteorológicas ao redor do mundo, devido ao elevado custo de aquisição dos sensores.

Diante da necessidade de informações climáticas sobre a UV solar, dados podem ser obtidos a partir de diferentes métodos de estimativa. Existem três métodos diferentes de estimativa da radiação solar UV normalmente utilizados: modelos físicos, modelos empíricos e as Técnicas de Aprendizado de Máquinas (TAM). Dentre os três métodos citados, os modelos empíricos são os mais utilizados devido à simplicidade e acurácia próxima à dos sensores de medição da UV (Foyo-Moreno *et al.*, 1999; Escobedo *et al.*, 2009; Villán *et al.*, 2010; Huang *et al.*, 2011). Ao contrário deles, os modelos físicos exigem a disponibilidade de dados de um número grande de variáveis meteorológicas, a fim de caracterizar de forma rigorosa o perfil vertical da atmosfera, sendo esse um fator que dificulta sua utilização. Em relação às TAM, o uso delas na estimativa da radiação solar foi foco de estudos nos últimos vinte anos com o avanço na tecnologia dos computadores e a disponibilidade de compilações gratuitas e com interface gráfica. A grande maioria desses estudos avaliou o uso das TAM na estimativa da R_G e verificaram maior acurácia delas em relação aos modelos empíricos, sendo a acurácia dependente do conjunto de variáveis de entrada utilizado na estimativa (Yadav e Chandel, 2014; Qazi *et al.*, 2015; Silva *et al.*, 2017). As TAM mais testadas foram as Redes Neurais Artificiais (RNA) e a Máquina de Vetores de Suporte (SVM) (Qazi *et al.*, 2015; Silva *et al.*, 2017; Meenal e Selvakumar, 2018; Quej *et al.*, 2018). E dentre as duas TAM citadas, a SVM é considerada o melhor método de estimativa da R_G , sendo capaz de gerar estimativas com maior acurácia utilizando um número menor de variáveis de entrada (Chen *et al.*, 2013;

Qazi *et al.*, 2015; Belaid; Mellit, 2016; Silva *et al.*, 2017; Quej *et al.*, 2018; Meenal e Selvakumar, 2018). Santos *et al.* (2016) confirmaram a superioridade da SVM frente a RNA e aos modelos empíricos clássicos também na estimativa da radiação direta horária.

Contudo, poucos trabalhos avaliaram o uso das TAM na estimativa da UV solar, a grande maioria com foco na partição diária e utilizando as RNA (Barbero *et al.*, 2006; Leal *et al.*, 2011; Jacovides *et al.*, 2015). Em tais trabalhos, a RNA apresentou o mesmo nível de acurácia dos modelos empíricos quando são utilizadas a R_G , o índice de transmissividade atmosférica K_T e a massa óptica relativa do ar como variáveis de entrada. Reduções expressivas nos erros de estimativa ocorrem mediante a adição de outras variáveis de entrada tais como a coluna total de ozônio, temperatura do ar, umidade relativa do ar, fotoperíodo e brilho solar (Jacovides *et al.*, 2015). Quando é considerado o uso da SVM na estimativa da UV solar, apenas um trabalho é verificado na literatura, o qual avaliou o desempenho da SVM na estimativa da UV solar diária em Botucatu/SP/Brasil testando cinco conjuntos de variáveis de entrada diferentes, sem realizar a comparação com outros métodos de estimativa (Almeida *et al.*, 2013). Segundo Almeida *et al.* (2013), as variáveis que permitiram maior acurácia nas estimativas da UV solar diária utilizando a SVM foram a radiação solar no topo da atmosfera, radiação solar ultravioleta no topo da atmosfera e R_G . No caso específico da utilização da SVM na estimativa da UV horária, testando diferentes conjuntos de variáveis de entrada a fim de selecionar a mais adequada e realizando a comparação com outros métodos de estimativa, nenhum estudo é verificado na literatura.

Com base no exposto acima, o objetivo do presente estudo foi determinar por meio de análise comparativa qual é o método de maior acurácia na estimativa da UV solar horária, levando em consideração o uso de modelos empíricos clássicos e de diferentes conjuntos de variáveis de entrada nas técnicas de aprendizado de máquinas RNA e SVM. Quatro modelos empíricos foram calibrados para a estimativa da UV solar horária em Botucatu/SP/Brasil, sendo três destes dependentes apenas da medição da radiação global (variável tipicamente medida na maioria das estações meteorológicas no Brasil). Os conjuntos de variáveis de entrada testados nas duas TAM foram equivalentes aos utilizados por cada um dos modelos empíricos ajustados, a fim de permitir que todos os métodos de estimativa testados operassem sob as mesmas condições iniciais.

2. Material e Métodos

2.1. Local de estudo e obtenção dos dados

O presente estudo é baseado em dados radiométricos horários registrados na Estação Radiométrica de Botucatu

(ERB), instalada na Faculdade de Ciências Agrárias da UNESP em Botucatu (22.85° S, 48.43° O, 742 m), durante oito anos (2000-2007). Botucatu é considerada uma cidade rural, com território de 1.482,87 km² e uma população de 127 mil habitantes (IBGE, 2016). Não há grandes indústrias no município, a frota de veículos automotivos é de aproximadamente 68 mil e a agricultura é diversificada. As principais culturas agrícolas em Botucatu são o eucalipto (252 km²), cana de açúcar (150 km²) e a laranja (115 km²) (IBGE, 2016).

O clima de Botucatu é caracterizado pelo inverno frio e seco (junho-agosto) e verão quente e úmido (dezembro-fevereiro), sendo classificado como Cwa pelo critério de classificação de Köppen. Durante o inverno, a temperatura média do ar é 17,6 °C e a umidade relativa do ar 67,3%, enquanto no verão elas são 22,7 °C e 78,2%, respectivamente (Escobedo *et al.*, 2009). O ciclo anual de chuvas é composto por dois períodos distintos: (i) período chuvoso e (ii) período seco. No período chuvoso (outubro-março) ocorre de 75 a 90% do volume total de precipitação acumulada anual, quando as chuvas e a elevada frequência de dias com condição de céu nebuloso são resultantes da formação da Zona de Convergência do Atlântico Sul (ZCAS) e de sistemas frontais (CPTEC, 2010; Teramoto e Escobedo, 2012). No período seco (abril-setembro), os valores mensais de precipitação acumulada permanecem abaixo de 100 mm e há a predominância de dias com condição de céu claro. Aumentos na concentração de aerossóis na atmosfera local podem ocorrer durante o período seco devido aos incêndios em áreas com vegetação nativa ou pastagem em outros estados do Brasil e países da América do Sul (Holben *et al.*, 2001; Freitas *et al.*, 2005; Mariano *et al.*, 2010).

As medições da radiação global (R_G) e da UV solar foram realizadas utilizando um piranômetro modelo PSP da Eppley e um radiômetro modelo CUV3 da Kipp & Zonen, respectivamente. As imprecisões estimadas nas medições do radiômetro CUV3 e do piranômetro PSP são de $\pm 5\%$ (Huang *et al.*, 2011) e de $\pm 4.1\%$ (Reda *et al.*, 2008), respectivamente. Ambos os sensores foram calibrados periodicamente de acordo com o método comparativo sugerido pela OMM (WMO, 2008). Na aquisição de dados foi utilizado um datalogger modelo CR23X da Campbell, operando na frequência de 1 Hz e armazenando (médias) a cada 5 minutos. Os dados medidos foram integrados na partição horária (expressa em MJm⁻² h⁻¹) utilizando um software apropriado (Chaves e Escobedo, 2000). A base de dados da estação radiométrica foi estendida com dados horários de transmissividade atmosférica K_T , UV extra-terrestre (UV_0), massa óptica relativa do ar (m_r) e coluna total de ozônio oblíqua (SOC). O K_T é adimensional e expresso como a razão entre a R_G e a radiação solar extra-terrestre (R_0). A R_0 e a UV_0 horárias foram calculadas a partir da Eq. (1) (Iqbal, 1983):

$$X_0 = I_x \left(\frac{12}{\pi} \right) E_0 \int_{\omega_1}^{\omega_2} \text{sen} \gamma \, d\omega \quad (1)$$

onde X_0 representa a R_0 ou a UV_0 ; γ é o ângulo de elevação solar; I_x é a constante solar (78,3 Wm⁻² para o espectro UV e 1.367,1 Wm⁻² para o espectro total) (Iqbal, 1983; Gueymard, 2004); ω é o ângulo horário (expresso em graus); E_0 o fator de correção da excentricidade da órbita terrestre. A m_r é adimensional foi calculada a partir da Eq. (2) (Kasten e Young, 1989):

$$m_r = \frac{1}{\left[\cos(\theta_z) + 0,050572(96,0795 - \theta_z)^{-1,6364} \right]} \quad (2)$$

onde θ_z é o ângulo zenital (expresso em graus) e obtido por meio da metodologia proposta por Iqbal (1983). A SOC é expressa em DU foi calculada a partir da Eq. (3) (Antón *et al.*, 2008):

$$SOC = \frac{TOC}{\cos(\theta_z)} \quad (3)$$

onde TOC é a coluna total de ozônio diária, expressa em DU. Os dados de TOC utilizados foram medidos pelos satélites Earth Probe (período de 2000-2004) e Aura (período de 2005-2007), ambas as bases de dados obtidas de imagens disponíveis em <http://mirador.gsfc.nasa.gov/>. Tais imagens estão disponíveis em pontos de grade no formato HDF (do inglês “Hierarchical Data Format”) com resolução espacial de 1,00 x 1,00 graus (Nunes e Mariano, 2015). Para obtenção desses dados, foram extraídos os valores pontuais de TOC da imagem raster relativos às coordenadas geográficas da ERB. O satélite Earth Probe mediu a TOC através do sensor TOMS (do inglês “Total Ozone Mapping Spectrometer”) e satélite Aura através do sensor OMI (do inglês “Ozone Monitoring Instrument”), ambos os sensores com imprecisão na medição da ordem de $\pm 2\%$ (McPeters *et al.*, 2008).

Em uma etapa seguinte, as bases de dados foram sincronizadas, sendo consideradas horas com registros úteis apenas as que apresentaram a medição simultânea da R_G , UV e SOC . Após a sincronização, os dados disponíveis correspondiam a 26.530 dados (78,8% dos dados iniciais). Por último, com o propósito de calibrar os modelos empíricos, treinar as duas TAM e validar todos os métodos de estimativa, a base total de dados foi dividida em duas partes: i - base de dados para ajuste dos modelos empíricos e treinamento das TAM; ii - base de dados de validação dos métodos de estimativa. A parte (i) foi composta pelos dados registrados nos anos de 2000, 2001, 2002, 2004 e 2006, correspondente a 62,1% dos dados disponíveis após a sincronização. A parte (ii) foi composta pelos dados registrados nos anos de 2003 (3.507 dados), 2005 (3.418 dados) e 2007 (3.133 dados), sendo a soma de todos os

dados correspondente a 37,9% dos dados disponíveis pós a sincronização.

2.2. Modelos empíricos

O desenvolvimento de modelos empíricos ocorre com base na correlação entre a UV solar e uma ou mais variáveis meteorológicas. Durante as últimas duas décadas, diferentes modelos empíricos foram propostos na literatura, os quais podem ser divididos em dois grupos principais: G1 - modelos que estimam a UV solar a partir da radiação global; G2 - modelos baseados nos índices de transmissividade atmosférica K_{TUV} e K_T . O K_{TUV} corresponde à razão entre a radiação solar UV incidente na superfície terrestre e a radiação solar UV extraterrestre. Os modelos G1 expressam a UV como um percentual da radiação global, são mais simples, porém apresentam como principal desvantagem a dependência local. Além disso, eles apresentam menor acurácia devido a radiação global ser influenciada pelo vapor de água e pelas nuvens de maneira diferente da UV (Foyo-Moreno *et al.*, 1999; Huang *et al.*, 2011; Wang *et al.*, 2015). Ao contrário deles, os modelos G2 correlacionam a transmissividade atmosférica da radiação UV (K_{TUV}) com o K_T eliminando as dependências local e astronômica.

No presente estudo, quatro modelos empíricos foram calibrados e validados com base na literatura (Tabela 1), sendo um G1 e os demais G2.

Em todos os modelos apresentados na Tabela 1, “ a ” é o parâmetro calibrado para o local de estudo. O modelo FM proposto por Foyo-Moreno *et al.* (1999) relaciona a UV solar horária com seus principais atenuadores na atmosfera: nuvens (por meio do K_T) e ar seco (por meio do m_r). A m_r serve de indicativo da espessura óptica do percurso percorrido pela radiação solar ao longo da atmosfera (Kasten e Young, 1989). Na equação, a UV_{claro} (expressa em $MJ\ m^{-2}\ h^{-1}$) é a UV solar horária em condições de céu claro, que é calculada a partir da Eq. (4).

$$UV_{claro} = y_0 + A_1 e^{\left(\frac{-m_r}{T_1}\right)} + A_2 e^{\left(\frac{-m_r}{T_2}\right)} \quad (4)$$

onde os coeficientes y_0 , A_1 , A_2 , T_1 e T_2 são obtidos da correlação entre a UV_{claro} medida e a m_r . O critério utilizado para classificar as condições de céu como claro foi o $K_T > 0,65$ (Escobedo *et al.*, 2009).

O modelo HM é uma variação do modelo FM. Nele, a estimativa da UV_{claro} é realizada a partir das variáveis de entrada m_r e SOC (Eq. (5)). Assim, o modelo HM con-

sidera não só a influência das nuvens e do ar seco sobre a UV solar, mas também a influência do ozônio atmosférico.

$$UV_{claro} = y_0 + A_1 e^{\left(\frac{-m_r}{T_1}\right)} + A_2 e^{\left(\frac{-m_r}{T_2}\right)} + A_3 e^{\left(\frac{-SOC}{T_3}\right)} + A_4 e^{\left(\frac{-m_r}{T_4}\right)} \quad (5)$$

onde os coeficientes y_0 , A_1 , A_2 , A_3 , A_4 , T_1 , T_2 , T_3 e T_4 são obtidos da correlação entre a UV_{claro} medida e os parâmetros m_r e SOC . Na Tabela 2 são apresentados os coeficientes de todos os modelos calibrados para Botucatu.

2.3. Técnicas de aprendizado de máquinas

As redes neurais artificiais (RNA) são estruturas compostas por nós computacionais (neurônios) com elevada capacidade de processamento de informações e habilidade de aprendizado e adaptabilidade. O treinamento ocorre através de um conjunto de dados (conjunto de treinamento) e de um programa de treinamento (algoritmo de aprendizado), sendo o processo análogo ao aprendizado biológico. O algoritmo de aprendizado carrega os casos exemplos do conjunto de treinamento e atribui pesos a eles a fim de reduzir os erros no momento da estimativa.

Dentre os tipos de redes neurais existentes, a Perceptron de Multicamadas (MLP) tem sido a mais utilizada na estimativa de dados de radiação solar (Barbero *et al.*, 2006; Leal *et al.*, 2011; Jacovides *et al.*, 2015). A MLP é composta pelas seguintes estruturas: camadas de neurônios artificiais, cada uma responsável pela recepção dos sinais (dados de entrada), análise dos sinais (camadas ocultas) e saída dos dados estimados (camada de saída); um algoritmo de aprendizado, correspondente a um conjunto de regras definidas para resolver falhas no aprendizado e minimizar os erros; e uma função de ativação, que define a influência do algoritmo sobre a ativação dos neurônios (Haykin, 1998). A função de ativação usada pela MLP é a típica de padrão sigmoide e o algoritmo de treinamento é o Levenberg-Marquardt de retro propagação (Marquardt, 1963; Rumelhart *et al.*, 1986).

A SVM foi originalmente desenvolvida por Vapnik (1999) com base na Teoria do Aprendizado Estatístico, sendo hoje amplamente utilizada para resolver problemas de classificação e regressão. Durante o processo de treinamento da SVM para a solução de problemas de regressão, é estimada uma função a partir do conjunto de exemplos $[(x_i, y_i)]_{i=1}^n$, onde $x_1 \in R^n$ é o vetor exemplo formado

Tabela 1 - Lista de modelos para estimativa da UV solar horária testados.

Autor do modelo	Representação	Variáveis requeridas	Equação
Ogunjobi e Kim (2004)	LM	R_G	$UV = a R_G$
Ogunjobi e Kim (2004)	LT	UV_0 e K_T	$UV = a UV_0 K_T$
Foyo-Moreno <i>et al.</i> (1999)	FM	K_T e m_r	$UV = a [K_T UV_{claro}]$
Huang <i>et al.</i> (2011)	HM	K_T , m_r e SOC	$UV = a [K_T UV_{claro}]$

Tabela 2 - Coeficientes dos modelos empíricos ajustados para Botucatu.

Modelo	Coeficiente ajustado									
	a	y_0	A_1	A_2	A_3	A_4	T_1	T_2	T_3	T_4
LM	0,04									
LT	0,76									
FM	1,50	0,01	1,01	0,11			0,40	2,58		
HM	1,49	0,02	2,0	0,24	0,41	0,02	0,24	0,98	78,14	394,87

pelas variáveis de entrada, $y_1 \in \{-1, +1\}$ é a variável de saída e n é o número total de exemplos presentes no conjunto de treinamento (Vapnik, 1999). Assim, a função f obtida no treinamento será capaz de receber um vetor x_i e fornecer uma predição y_i . O desempenho e capacidade de aprendizado da SVM na regressão são atribuídos ao uso do conjunto de funções de Kernel, as quais diagramam a informação para um espaço dimensional elevado (Vapnik, 1999). A SVM fornece 4 funções de Kernel básicas: sigmoide, linear, polinomial e função de base radial (RBF). A RBF é a função de Kernel mais utilizada (Mohammadi et al., 2015). Para permitir maior flexibilidade na aplicação da RBF, é necessário o ajuste adequado dos parâmetros C (custo), γ (gamma) e ε (epsilon) (Vapnik, 1999). Maiores detalhes sobre a teoria da SVM podem ser obtidos em Vapnik (1999).

As versões da RNA e SVM utilizadas no presente trabalho são as implementadas e disponíveis no software WEKA (Hall et al., 2009; Witten et al., 2011). Baseado nos modelos empíricos, os modelos RNA e SVM testados utilizaram como variáveis de entrada a R_G , UV_0 , K_T , m_r e SOC . Nesse caso, foram avaliados quatro modelos RNA e quatro SVM (Tabela 3), que serão simbolizados como RNA(x) ou SVM(x), onde o x indica o agrupamento de variáveis de entrada utilizado. Devido à eficiência computacional, no presente estudo foi utilizada a função de Kernel de Base Radial (RBF) na SVM.

2.4. Validação dos métodos de estimativa

A avaliação do desempenho dos métodos de estimativa testados foi realizada com base nos parâmetros estatísticos amplamente utilizados Erro Padrão Médio (MBE) e Erro Médio Absoluto (MAE), ambos expressos em valores percentuais (Willmott e Matsuura, 2005). O MBE é usado para mostrar as tendências de superestimação ($MBE > 0$) ou subestimação ($MBE < 0$) dos métodos de

estimativa e o MAE expressa a extensão do erro. Optou-se pelo uso do MAE e não do RMSE (Raiz Quadrada do Erro Médio) levando em consideração o estudo divulgado por Willmott e Matsuura (2005), o qual comprovou a superioridade do MAE frente ao RMSE como indicador da extensão do erro nas estimativas. O MBE e o MAE foram calculados a partir das seguintes Equações:

$$MBE = 100 \frac{n^{-1} \sum_{i=1}^n (UV_{E_i} - UV_{M_i})}{\overline{UV}_M} \quad (5)$$

$$MAE = 100 \frac{n^{-1} \sum_{i=1}^n |UV_{E_i} - UV_{M_i}|}{\overline{UV}_M} \quad (6)$$

onde UV_E é o valor estimado; UV_M é o valor medido; \overline{UV}_M é a média dos valores medidos e n é o número de observações.

A base de dados utilizada na calibração foi composta pelos registros feitos ao longo de um período de três anos (2005- 2007), sendo cada um dos anos considerado uma repetição. Assim, na validação de cada um dos métodos de estimativa foram obtidos três valores de cada um dos dois parâmetros estatísticos (MBE e MAE), os quais foram utilizados no cálculo das médias. A análise de variância (ANOVA) seguida do teste F ($p < 5\%$) junto do teste de Tukey ($p = 5\%$) foi realizado para a comparação das médias do MAE. O software utilizado para a realização dos testes estatísticos foi o SISVAR (Ferreira, 1998). Na comparação das médias, considerou-se o delineamento experimental fatorial 3 x 4 (três métodos combinados com quatro conjuntos de variáveis de entrada) e cada ano nos dados de validação foi considerado uma repetição. Esses procedimentos são realizados com o objetivo de verificar se as diferenças na acurácia entre os métodos testados são estatisticamente significativas.

3. Resultados e Discussão

Na Tabela 4 são apresentados os valores dos parâmetros estatísticos MBE e MAE obtidos na validação dos modelos empíricos calibrados. Na validação, foram utilizados os dados medidos no período de 2003, 2005 e 2007 (3 anos), sendo cada um desses anos considerado uma repetição para o cálculo das médias de MBE e MAE. Os valores horários de UV estimados por cada modelo são

Tabela 3 - Conjuntos de variáveis de entrada testados na RNA e na SVM.

Modelo	Variáveis de entrada	
RNA(1)	SVM(1)	R_G
RNA(2)	SVM(2)	UV_0 e K_T
RNA(3)	SVM(3)	UV_0 , K_T e m_r
RNA(4)	SVM(4)	UV_0 , K_T , m_r e SOC

apresentados na Fig. 1 na forma de gráfico de dispersão (estimados versus medidos) para comparação visual. A regressão linear entre os valores estimados e os medidos foi calculada para todos os modelos (linha cinza contínua) e foram apresentadas junto aos coeficientes de determinação obtidos. A linha traço ponto negra representa a reta 1:1.

Como esperado, a acurácia dos modelos empíricos aumentou acompanhando o aumento no número de variáveis de entrada utilizadas, do LM até o HM, com tendência de diminuição nos valores do MBE e do MAE. O mesmo resultado foi observado por Huang *et al.* (2011) em estudo similar realizado em duas localidades na bacia do rio Yangtze, China. Porém, ressalta-se que as variações observadas aqui nos valores dos parâmetros estatísticos foram pequenas. No modelo LM verificou-se uma tendência de superestimativa de 5,39%, com extensão dos erros nas estimativas de 9,51%. No HM também houve tendência de superestimativa de 4,07%, com extensão dos erros nas estimativas de 8,95%. A menor acurácia apresentada pelo LM em relação aos outros modelos ocorre devido ao uso da R_G como única variável de entrada, a qual é influenciada pelo vapor de água e pelas nuvens de maneira diferente da UV (Wang *et al.*, 2015). Em todos os modelos o coeficiente de determinação apresentou valores próximos de 0,97 (Fig. 1), ou seja, os níveis de concordância entre os valores medidos e os estimados foram próximos de 97%.

Considerando agora o desempenho das técnicas de aprendizado de máquinas RNA e SVM na estimativa da UV solar horária, na Tabela 5 são apresentados os valores de MBE e MAE obtidos na validação. Na Fig. 2 são apresentados os gráficos de dispersão entre os valores estimados e os medidos na ERB para comparação visual. A regressão linear entre os valores estimados e os medidos foi calculada para todos os modelos (linha cinza contínua) e foram apresentadas junto aos coeficientes de determinação obtidos. A linha traço ponto negra representa a reta 1:1.

Os valores do coeficiente de determinação variaram de 0,97 no modelo RNA(1) a 0,99 no modelo RNA(4). Os níveis de acurácia da RNA foram baixos quando foram utilizados os conjuntos de variáveis de entrada $\{R_G\}$ ou $\{K_T, UV_0\}$, os quais fizeram o desempenho dela ser inferior até em relação aos modelos empíricos equivalentes (LM e LT, respectivamente). A partir da adição da m_r no conjunto de variáveis de entrada $\{K_T, UV_0\}$, houve um

Tabela 4 - Valores de MBE e MAE e desvios padrão obtidos na validação dos modelos empíricos calibrados para a estimativa da UV horária em Botucatu.

Modelo	MBE(%)	MAE(%)
LM	5,39 ± 4,53	9,51 ± 0,36
LT	5,18 ± 4,58	9,40 ± 0,58
FM	4,05 ± 4,58	9,26 ± 0,20
HM	4,07 ± 4,51	8,95 ± 0,01

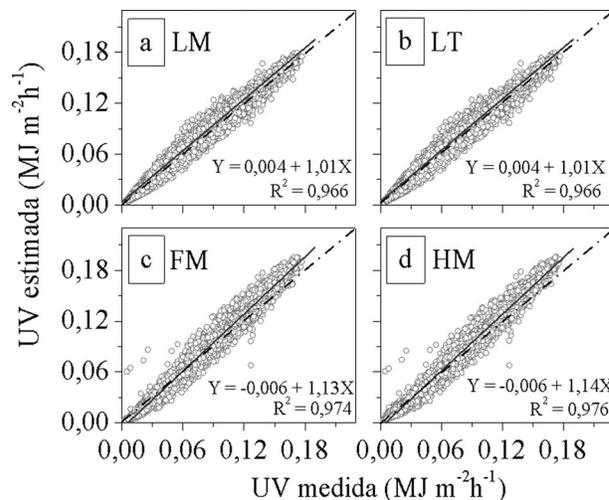


Figura 1 - Comparação entre valores de UV solar horária estimados pelos modelos empíricos e os medidos na ERB.

acentuado aumento no nível de acurácia da RNA na estimativa da UV solar horária, com a diminuição do MAE de 10,64% para 4,48%. Tal resultado está de acordo com o apresentado por Barbero *et al.* (2006), que através de análise bayesiana identificou a m_r como sendo uma variável de entrada importante na estimativa da UV solar utilizando RNA. Ao compararmos os modelos RNA(3) e RNA(4) entre si, a diferença observada entre os valores de MAE foi muito baixa, o que evidencia a baixa relevância da SOC como variável de entrada na RNA para estimativa da UV solar horária. Em relação aos valores de MBE, a RNA(3) apresentou uma tendência de subestimativa de 0,37% e a RNA(4) uma tendência de superestimativa de 1,46%. Porém, quando os modelos RNA(3) e RNA(4) são comparados aos seus modelos empíricos equivalentes em relação ao conjunto de variáveis de entrada utilizados (FM e HM, respectivamente), verifica-se uma acentuada diferença entre os valores de MAE (~50% menor nos modelos RNA).

Quando comparada aos modelos empíricos ou a RNA, a SVM foi capaz de identificar melhor a correlação entre a UV solar horária e os parâmetros atmosféricos testados. Na maioria dos casos, os valores estimados utilizando a SVM apresentaram níveis de concordância com os valores medidos equivalentes a 99% (Fig. 2 f, g, h). No geral, os erros nas estimativas gerados pela SVM foram baixos quando comparados aos erros gerados pelos modelos empíricos e pela RNA, com extensão da SVM(1), que apresentou nível de acurácia similar ao observado no LM. Um detalhe importante observado foi o nível elevado de acurácia atingido pela SVM ao utilizar o conjunto de variáveis de entrada $\{K_T, UV_0\}$, sendo superior ao obtido pelos modelos empíricos e RNA utilizando as mesmas variáveis. Esse resultado está de acordo com o divulgado por outros autores em estudos realizados com a radiação global, os quais verificaram que a SVM é capaz de atingir

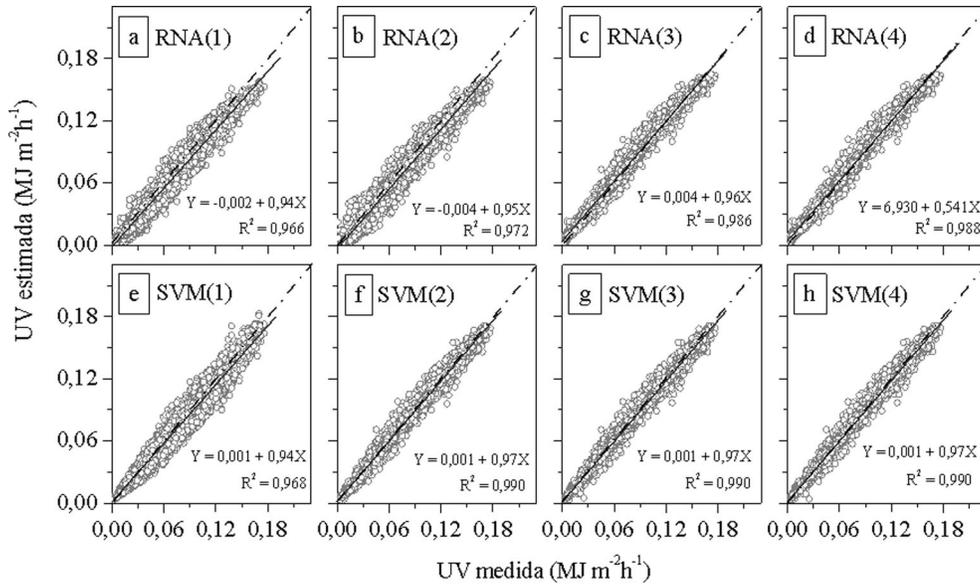


Figura 2 - Comparação entre valores de UV solar horária estimados através da RNA e da SVM e os medidos na ERB.

um nível de acurácia maior utilizando um número menor de variáveis de entrada, quando é comparada aos modelos empíricos ou à RNA (Qazi *et al.*, 2015; Belaid; Mellit, 2016; Silva *et al.*, 2017; Quej *et al.*, 2018; Selvakumar, 2018). Similar ao observado na RNA, na SVM a adição do SOC ao conjunto de variáveis de entrada não trouxe aumento expressivo na acurácia.

A fim de verificarmos agora se as diferenças no desempenho observadas são significativas entre os métodos de estimativa da UV solar testados e em relação aos conjuntos de variáveis de entrada utilizados, na Tabela 6 são apresentados os resultados da ANOVA e do teste F ($p < 5\%$) junto do teste de Tukey ($p = 5\%$) para comparação das médias de MAE (extensão dos erros gerados na estimativa).

Letras maiúsculas indicam o método utilizado e letras minúsculas o conjunto de variáveis de entrada. Médias seguidas da mesma letra, maiúscula na coluna e minúscula

na linha, não se diferem entre si, pelo teste de Tukey, a 5% de probabilidade ($n = 3$).

Entre os quatro modelos empíricos testados, as diferenças de acurácia, considerando a extensão dos erros gerados na estimativa, não foram significativas estatisticamente (Tabela 6, segunda coluna). Em relação à RNA, a RNA(3) apresentou a melhor acurácia e o uso da SOC como variável de entrada não trouxe melhorias significativas no desempenho. Quando comparada à RNA, a SVM atingiu o maior nível de acurácia utilizando um número menor de variáveis de entrada $\{UV_0; K_T\}$. A RNA e a SVM apresentaram o mesmo nível de acurácia que todos os modelos empíricos testados quando utilizaram apenas a R_G como variável de entrada (Tabela 6, segunda linha). Os erros gerados na estimativa da UV solar horária foram similares estatisticamente entre os seguintes métodos: RNA(3), RNA(4), SVM(2), SVM(3) e SVM(4).

4. Conclusões

Os resultados das validações e do estudo comparativo indicaram que a SVM possui maior acurácia quando comparada aos modelos empíricos e à RNA, sendo ela capaz de gerar melhores resultados utilizando um número menor de variáveis de entrada. Dentre os métodos testa-

Tabela 5 - Valores de MBE e MAE e desvios padrão observados na estimativa da UV horária em Botucatu através da RNA e da SVM utilizando diferentes conjuntos de variáveis de entrada.

Modelo	MBE (%)	MAE (%)
RNA(1)	$-8,07 \pm 3,99$	$11,07 \pm 2,29$
RNA(2)	$-7,29 \pm 2,03$	$10,64 \pm 3,42$
RNA(3)	$-0,37 \pm 1,22$	$4,48 \pm 1,27$
RNA(4)	$1,46 \pm 0,48$	$4,54 \pm 0,94$
SVM(1)	$-3,70 \pm 1,25$	$8,80 \pm 0,84$
SVM(2)	$-1,25 \pm 0,79$	$4,18 \pm 1,16$
SVM(3)	$-1,30 \pm 0,91$	$4,06 \pm 0,92$
SVM(4)	$-1,17 \pm 0,76$	$3,87 \pm 1,11$

Tabela 6 - Desdobramento das interações entre o conjunto de variáveis de entrada e o método de estimação.

Conjunto de variáveis de entrada	Modelo empírico	RNA	SVM
R_G	9,51 Aa	11,07 Aa	8,80 Aa
UV_0 e K_T	9,40 Aa	10,64 Aa	4,18 Bb
UV_0, K_T e m_r	9,26 Aa	4,48 Bb	4,06 Bb
UV_0, K_T, m_r e SOC	8,95 Aa	4,54 Bb	3,87 Bb

dos, a SVM utilizando o conjunto de variáveis de entrada $\{UV_0, K_T\}$ é considerada a melhor alternativa devido ao menor número de variáveis de entrada e relativa acurácia. Com uso da SVM e tendo a necessidade da medição apenas da radiação solar global, é possível estimar a UV solar horária com tendência de subestimativa de 1,25% e extensão do erro de 4,18%. A adição da coluna total de ozônio medida pelo satélite não é capaz de reduzir significativamente os erros na estimativa da UV solar horária através da SVM ou da RNA. Todos os modelos empíricos testados no presente estudo não se diferenciam significativamente entre si em relação à acurácia.

Agradecimentos

À CAPES, CNPq e FAPESP pelo apoio financeiro e à NASA pelos dados meteorológicos utilizados nesta pesquisa.

Referências Bibliográficas

- ALMEIDA, T.N.S.; TERAMOTO, E.T.; DAL PAI, A.; RODRIGUES, D.; ESCOBEDO, J.F. Estimativa da radiação solar ultravioleta em Botucatu/SP/Brasil utilizando Máquina de Vetores de Suporte (SVM). **Tekne e Logos**, v. 4, n. 1, p. 183-193, 2013.
- ANTÓN, M.; SERRANO, A.; CANCELLO, M.L.; GARCÍA, J.A. A multiple regression analysis between UV radiation measurements at Badajoz and ozone, reflectivity and aerosols estimated by TOMS. **Physica Scripta**, v. 118, p. 21-23, 2008.
- BARBERO, F.J.; LÓPEZ, G.; BATLLES, F.J. Determination of daily solar ultraviolet radiation using statistical models and artificial neural networks. **Annales Geophysicae**, v. 24, p. 2105-2116, 2006.
- 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**, v. 118, p. 105-118, 2016.
- CHAVES, M.; ESCOBEDO, J.F. A software to process daily solar radiation data. **Renewable Energy**, v. 19, p. 339-5, 2000.
- CHEN, J.L.; LI, G.S.; WU, S.J. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. **Energy Conversion and Management**, v. 75, p. 311-318, 2013.
- CPTEC - Centro de Previsão do Tempo e Estudos Climáticos/INPE. Climanálise - **Boletim de monitoramento e análise climática**. Disponível em: <http://climanalise.cptec.inpe.br/~rclimanl/boletim/>. Acesso em: maio 2014.
- ESCOBEDO, J.F.; GOMES, E.N.; OLIVEIRA, A.P.; SOARES, J. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. **Applied Energy**, v. 86, p. 299-10, 2009.
- ESCOBEDO, J.F.; GOMES, E.N.; OLIVEIRA, A.P.; SOARES, J. 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.
- FERREIRA, D.F. **Sisvar - sistema de análise de variância para dados balanceados**. Lavras: UFLA, 1998.
- FOYO-MORENO, I.; VIDA, J.; ALADOS-ARBOLEDAS, L. A simple all weather model to estimate ultraviolet solar radiation (290-385 nm). **Journal of Applied Meteorology**, v. 38, p. 1021-1026, 1999.
- FREITAS, S.R.K.; LONGO, M.A.; DIAS, P.L.S.; DIAS, R.S.; CHATFIELD, E.; et al. Monitoring the Transport of Biomass Burning Emissions in South America. **Environmental Fluid Mechanics**, v. 5, p. 135-167, 2005.
- GUEYMARD, C. The sun's total and spectral irradiance for solar energy applications and solar radiation models. **Solar Energy**, v. 76, p. 423-453, 2004.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMRNA, P.; et al. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, p. 10-18, 2009.
- HAYKIN, S. **Neural networks: A comprehensive foundation: 2th edition**. Upper Sanddle River: Prentice Hall, 1998.
- HIDEG, E.; JANSEN, M.A.; STRID, A. UV-B exposure, ROS, and stress: inseparable companions or loosely linked associates? **Trends Plant. Sci.**, v. 18, p. 107-115, 2013.
- HOLBEN, B.N.D.; TANRÉ, D.; SMIRNOV, S.; ECK, T.; SLUTSKER, I.; et al. An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. **Journal of Geophysical Research**, v. 106, p. 12067-12097, 2001.
- HUANG, M.; JIANG, H.; JU, W.; XIAO, Z. Ultraviolet radiation over two lakes in the Middle and lower reaches of the Yangtze River, China: an innovative model for UV estimation. **Terrestrial, Atmospheric and Oceanic Sciences**, v. 22, n. 5, p. 491-506, 2011.
- Instituto Brasileiro de Geografia e Estatística Cidades - IBGE. **Senso dos municípios paulistas**. Disponível em: <http://cidades.ibge.gov.br/xtras/uf.php?lang=&coduf=35&search=sao-paulo>. Acesso em: out. 2016.
- IQBAL, M. An introduction to solar radiation. Londres: Academic Press, 1983.
- JACOVIDES, C.P.; TYMVIOS, F.S.; BOLAND, J.; TSITOURI, M. Artificial Neural Network models for estimating daily solar global UV, PAR and broadband radiant fluxes in an eastern Mediterranean site. **Atmospheric Research**, v. 152, p. 138-145, 2015.
- JACOVIDES, C.P.; ASSIMAKOPOULOS, V.D.; TYMVIOSA, F.S.; THEOPHILOU, K.; ASIMAKOPOULOS, D.N. Solar global UV (280-380 nm) radiation and its relationship with solar global radiation measured on the island of Cyprus. **Energy**, v. 31, p. 2728-2738, 2006.
- KASTEN, F.; YOUNG, A.T. Revised optical air mass tables and approximation formula. **Applied Optics**, v. 28, n. 22, p. 4735-4738, 1989.
- KATARIA, S.; JAJOO, A.; GURUPRASAD, K.N. Impact of increasing Ultraviolet-B (UV-B) radiation on photosynthetic processes. **Journal of Photochemistry and Photobiology B: Biology**, v. 137, p. 55-66, 2014.
- LEAL, S.S.; TIBA, C.; PIACENTINI, R. Daily UV radiation modeling with the usage of statistical correlations and artificial neural networks. **Renewable Energy**, v. 36, p. 3337-3344, 2011.
- LIPINSKI, V.M.; SANTOS, T.G.D.; SCHUCH, A.P. An UV-sensitive anuran species as an indicator of environmental quality of the Southern Atlantic Rainforest. **Journal of**

- Photochemistry and Photobiology B: Biology**, v. 165, p. 174-181, 2016.
- LIU, B.; LIU, X.B.; LI, Y.S.; HEBERT, S.J. Effects of enhanced UV-B radiation on seed growth characteristics and yield components in soybean. **Field Crops Research**, v. 154, p. 158-163, 2013.
- MARIANO, G.L.; LOPES, R.J.S; JORGE, M.P.P.M.; LAND-ULFO, E. Assessment of biomass burnings activity with the synergy of sun photometric and LIDAR measurements in São Paulo, Brazil. **Atmospheric Research**, v. 98, p. 486-499, 2010.
- MARQUARDT, D. An algorithm for least-squares estimation of nonlinear parameters. **Journal of Applied Mech.**; v. 11, p. 431-441, 1963.
- MCPETERS, R.; KROON, M.; LABOW, G.; BRINKSMA, E.; BALIS, D.; *et al.* Validation of the Aura ozone monitoring instrument total column ozone product. **Journal of Geophysical Research**, v. 113, p. 1-9, 2008.
- MEENAL, R.; SELVAKUMAR, I. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. **Renewable Energy**, v. 121, p. 324-343, 2018.
- MOHAMMADI, K.; SHAMSHIRBAND, S.; ANISI, M.H.; ALAM, K.A.; PETKOVIC, D. Support vector regression based prediction of global solar radiation on a horizontal surface. **Energy Conversion and Management**, v. 91, p. 433-441, 2015.
- NUNES, A.B.; MARIANO, G.L. **Meteorologia em tópicos: Volume 2**. Pelotas: Clube dos Autores, 2015.
- OGUNJOBI, K.O.; KIM, Y.J. Ultraviolet (0.280-0.400 μm) and broadband solar hourly radiation at Kwangju, South Korea: Analysis of their correlation with aerosol optical depth and clearness index. **Atmospheric Research**, v. 71, p. 193-214, 2004.
- QAZI, A.; FAYAZ, H.; WADI, A.; RAJ, R.G.; RAHIM, N.A.; *et al.* 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, 2015.
- QUEJ, V.H.; ALMOROX, J.; ARNALDO, J.A.; SAITO, L. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. **Journal of Atmospheric and Solar Terrestrial Physics**, v. 155, p. 62-70, 2018.
- REDA, I.M.; MYERS, D.R.; STOFFEL, T.L. Uncertainty Estimate for the Outdoor Calibration of Solar Pyranometers: A Metrologist Perspective. NCSLI Measure. **Journal of Measurement Science**, v. 3, p. 58-66, 2008.
- RUMELHART, D.E.; HINTON G.E.; WILLIAMS, R.J. Learning representations by back-propagation errors. **Nature**, v. 323, p. 533-536, 1986.
- SANTOS, C.M.; ESCOBEDO, J.F.; TERAMOTO, E.T.; SILVA, S.H.M.G. Assessment of ANN and SVM models for estimating normal direct irradiation (H_b). **Energy Conversion and Management**, v. 126, p. 826-836, 2016.
- SILVA, M.B.P.; ESCOBEDO, J.F.; SANTOS, C.M.; ROSSI, T.J.; GORLA-SILVA, S.H.M. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN Techniques to estimate the daily Global Solar Irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar-Terrestrial Physics**. v. 155, p. 62-70, 2017.
- SUCRÉ, S.; VIDUSSI, F.; MOSTAJIR, B.; CHARMANTIER, G.; LORIN-NEBEL, C. Impact of Ultraviolet-B radiation on planktonic fish larvae: Alteration of the osmoregulatory function. **Aquatic Toxicology**, v. 109, p. 194-201, 2012.
- TERAMOTO, E.T.; ESCOBEDO, J.F. Análise da frequência anual das condições de céu em Botucatu, São Paulo. **Revista Brasileira Engenharia Agrícola Ambiental**, v. 16, n. 9, p. 985-992, 2012.
- VAPNIK, V. N. An Overview of Statistical Learning Theory. **IEEE Transactions on Neural Networks**, v. 10, n. 5, p. 988-999, 1999.
- VILLÁN, M.D.; CASTRILLO, A.M.; SANTOS, J.B. Statistical models of UV total radiation and cloud effect study. **International Journal of Climatology**, v. 30, 1407-1413, 2010.
- WANG, F.; ZHEN, Z.; MIA, Z.; SUN, H.; SU, S.; *et al.* Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting. **Energy and Buildings**, v. 86, p. 427-438, 2015.
- WILLMOTT, C.J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climatic Research**, v. 30, p. 79-82, 2005.
- WITTEN, I.H.; FRANK, E.; HALL, M.A. **Data mining: practical machine learning tools and techniques: 3th edition**. San Francisco: Morgan Kaufmann, 2011.
- World Meteorological Organization - WMO. **Guide of Meteorological Instruments and Methods of Observation**. Geneva: World Meteorological Organization, 2008.
- YADAV, A.K.; CHANDEL, S.S. Solar radiation prediction using Artificial Neural Network techniques: A review. **Renewable and Sustainable Energy Reviews**, v. 33, p. 772-781, 2014.