

ESTIMATIVA DA FRAÇÃO DIFUSA DA IRRADIAÇÃO GLOBAL COM TÉCNICAS DE APRENDIZAGEM DE MÁQUINAS.

Edson Luis Bassetto – bassetto@utfpr.edu.br

UTFPR/Cornélio Procópio/PR, Departamento de Engenharia Elétrica

João Francisco Escobedo – escobedo@fca.unesp.br

UNESP/FCA/Botucatu/SP, Departamento de Engenharia Rural

Alexandre Dal Pai – adalpai@fca.unesp.br

UNESP/FCA/Botucatu/SP, Departamento de Bioprocessos e Biotecnologia

Resumo. Foram desenvolvidos modelos de estimativa da fração difusa (K_d) em função da fração transmitida da irradiação global (K_t) na partição horária sendo: Modelo Estatístico (**ME**); Redes Neurais Artificiais com Função de Base Radial (**RBF**); e Sistema Adaptativo de Inferência Neuro Fuzzy (**ANFIS**). O modelo estatístico utiliza como referência somente K_t e as técnicas uma combinação de seis (06) variáveis astronômicas, geográficas e meteorológicas. Os modelos utilizam uma base de sete anos (2000-2006) de medidas realizadas na Estação de Radiometria Solar de Botucatu/SP na partição horária, sendo parte para treinamento e outra para validação dividida em Ano típico (**AT**) e atípico (**AAT**). A equação do modelo estatístico gerada por regressão polinomial de 4^o ordem, apresenta coeficiente de determinação $R^2 = 0.80$ e na comparação dos valores medidos e estimados na validação, um coeficiente de correlação para ano típico (**AT**) $r_{AT}=0.90$ e para o ano atípico (**AAT**) de $r_{AAT}=0.89$, erro quadrático médio $rRMSE_{AT} = 30.55\%$ e $rRMSE_{AAT} = 27.97\%$. No desempenho das técnicas **RBF** e **ANFIS**, os modelos mostraram-se satisfatórios a partir da segunda combinação sendo para **RBF2** um coeficiente $r_{AT}=0.91$ e $r_{AAT}=0.90$ e erro de $rRMSE_{AT} = 29.63\%$ e $rRMSE_{AAT} = 26.93\%$ e para **ANFIS2** um $r_{AT}=0.93$ e $r_{AAT}=0.93$ com erro $rRMSE_{AT} = 25.13\%$ e $rRMSE_{AAT} = 22.76\%$. Para sexta combinação, a rede **RBF6** um coeficiente de $r_{AT}=0.92$ e $r_{AAT}=0.92$ e erro de $rRMSE_{AT} = 26.48\%$ e $rRMSE_{AAT} = 24.69\%$ e para **ANFIS6** um coeficiente $r_{AT}=0.95$ e $r_{AAT}=0.94$ e erro de $rRMSE_{AT} = 22.63\%$ e $rRMSE_{AAT} = 21.19\%$. Os indicadores mostram que as técnicas de aprendizagem de máquinas comparadas com modelo estatístico apresentaram um desempenho melhor com redução nos indicadores na ordem de 16% para rede **RBF6** e 34% para rede **ANFIS6** do erro quadrático médio para duas bases de validação (**AT** e **AAT**) em relação ao modelo **ME**.

Palavras-chave: Irradiação difusa, Redes neurais artificiais, Sistema Adaptativo de Inferência Neuro Fuzzy.

1. INTRODUÇÃO

A preocupação com fontes não renováveis de energia tem levado muitos pesquisadores a intensificar nos últimos anos o estudo com outras formas de energia. Entre elas, a energia solar tem se destacado por ser uma energia disponível, limpa e utilizada em diversas áreas e projetos, como na agronomia, no desenvolvimento agrícola ou em projetos de geração de energia através de conversores térmicos ou fotovoltaicos. Para isso é importante conhecer a eficiência e a disponibilidade desta energia ao longo do tempo, seja através de séries temporais ou modelos de estimativa, permitindo assim verificar a viabilização desses projetos.

Entre as radiações necessárias ao desenvolvimento destas áreas e projetos, a radiação difusa se destaca, pois tem influência em vários aspectos, principalmente os relacionados às suas características e sua influência em outros componentes. Apesar dessa importância, as medidas dessa radiação não são coletadas sistematicamente nas estações meteorológicas e as que são efetuadas necessitam de correções, tendo em vista as características físicas das formas de medidas e que podem comprometer tanto as medidas realizadas como suas projeções (Bortoloini et al., 2013; Kambezidis et al., 2017; Khorasanizadeh et al., 2016; Jamil e Siddiqui, 2017).

Quanto as formas de medidas da irradiação difusa, destacam-se a forma indireta que é realizada pela diferença das medidas realizadas da irradiação global e da irradiação direta e a forma direta que podem ser realizadas através do método com anel ou disco de sombreamento (Iqbal, 1983), sendo uma alternativa a montagem proposta por Melo e Escobedo (1994) que utiliza o método com anel de sombreamento, pela sua simplicidade e custo. Nesta montagem, o anel é fixo a um eixo no plano horizontal e inclinado na latitude local, sombreando um piranômetro que translada mecanicamente sobre esta superfície. Os outros métodos foram desenvolvidos por Drummond (1956) e Robinson e Stoch (1964), mas necessitam de equipamentos auxiliares como os rastreadores solares.

Como descrito, as medidas da irradiação difusa são restritas em algumas estações, institutos de pesquisas e centros meteorológicos, ao contrário de outras variáveis como a irradiação global, temperatura e umidade que é comumente medida nesses locais. Diante disto, a estimativa da irradiação difusa através de modelos e suas projeções pode suprir a ausência e as dificuldades de medidas.

Na literatura são propostos modelos para estimar a irradiação difusa sendo um dos primeiros trabalhos que relacionaram a fração transmitida da irradiação difusa (K_d) com a fração transmitida da irradiação global (K_t) para um determinado dia e local foi apresentado por Liu e Jordan (1960). Outros autores relacionaram também a fração da irradiação difusa com outros parâmetros como brilho solar, ângulo de elevação solar, declinação solar entre outros. Além

disso outros seguiram a mesma tendência nessa correlação utilizando outras partições de tempo na estimativa, como também ajustando as curvas geradas por regressão, por modelos lineares e polinomiais com uma ou mais variável na modelagem (Orgill e Hollands, 1977; Collares-Pereira e Rabl, 1979; Boland et al., 2001; Oliveira et al., 2002; Jacovides et al., 2006; Souza et al., 2011; Furlan et al., 2012; Kuo et al., 2014).

Apesar das técnicas de modelagem tradicional para estimativa da radiação solar, outras técnicas chamadas de Aprendizagem de Máquinas (AM) vem sendo utilizadas como as Redes Neurais Artificiais (RNA), Sistema Inferência Fuzzy (FIS), Máquinas de Vetores de Suporte (SVM) entre outras. Estas técnicas consistem de um conjunto de processos capazes de aprender e adquirir conhecimento acerca de um problema a ser solucionado. Dessas as mais utilizadas são as RNA's com arquiteturas do tipo *Perceptron* Multicamadas (MLP), Rede de Função de Base Radial (RBF), Regressão Generalizada (GRNN) e outras de forma híbrida como a ANFIS que utiliza RNA e FIS chamada de Sistema Adaptativo de Inferência Neuro Fuzzy (ANFIS). Essas técnicas possibilitam usar diversas variáveis de grandezas diferentes e disponíveis, como entrada ao modelo e combiná-las a fim de obter resultados melhores e satisfatórios para a estimativa. Conforme a literatura, algumas técnicas mostram-se melhores comparadas aos modelos estatísticos tradicionais na estimativa da irradiação solar (Soares et al., 2004; Elminir et al., 2007; Jiang, 2008; Alam et al., 2009; Behrang et al., 2010; Benghanem e Mellit, 2010; Quenj et al., 2012; Boata et al., 2012; Kisi, 2014; Khatib, 2015; Kumar et al., 2015; Azimi et al., 2016; Santos et al., 2016; Shamshirband et al., 2016).

Considerando o descrito, é proposto estimar a fração difusa (K_d) da irradiação global utilizando três técnicas, o modelo tradicional, a técnica de RNA do tipo RBF e ANFIS.

2. TÉCNICAS DE REDE NEURAL RBF, ANFIS E ESTATÍSTICO

Rede de função de base radial (RBF). As RNA são técnicas computacionais que apresentam um modelo matemático complexo e que possui uma estrutura inspirada em um neurônio humano e que adquirem conhecimento e habilidade de aprender através da experiência (Silva et al., 2010, Haykin, 2001), além de poder trabalhar com diversas variáveis de entrada em paralelo, sem o conhecimento prévio do evento ou modelo matemático a que se procura, como exemplo, estimar ou classificar dados ou informações. Entre algumas, as redes do tipo Função de Base Radial, conhecidas como RBF podem ser aplicadas em quase todos os tipos de problemas pela qual outras técnicas de RNA são utilizadas como a *Perceptron* Multicamadas (MLP) e Regressão Generalizada (GRNN), mas diferentes em suas estruturas, como no número de camadas e na quantidade de neurônios desta camada. As redes RBF são compostas por apenas uma camada intermediária além da camada de entrada e saída e com somente uma função de ativação conforme a Fig.1.

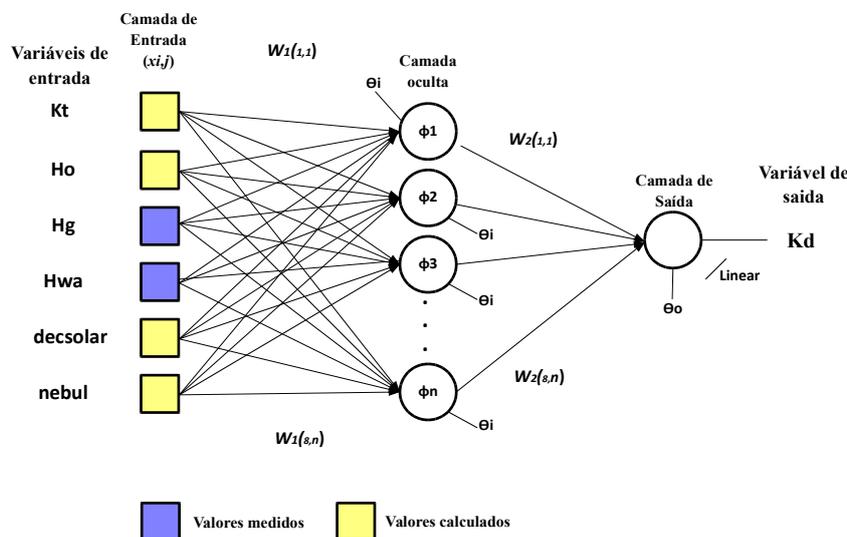


Figura 1: Arquitetura da Rede Neural de Base Radial. Adaptado de Silva et al, 2010.

A diferença no uso das redes RBF está na estratégia de treinamento utilizada, nos ajustes dos pesos de suas camadas, a intermediária, além da função de ativação, caracterizada por uma função de base radial como as gaussianas (Haykin, 2001). A arquitetura das redes RBF também são do tipo *feedforward* (alimentação para frente) com treinamento supervisionado, onde cada uma das entradas (x_n) se propaga pela camada intermediária em direção a saída da rede, porém seu treinamento difere de outras redes, pois é constituído de apenas duas fases bem distintos entre si.

A primeira fase está associada aos ajustes dos pesos dos neurônios da camada intermediária e não supervisionado, sendo apenas dependente da entrada, pois este ajuste está relacionado com a função de base radial, ou seja, com a alocação da função. Neste estágio, tem-se a determinação dos parâmetros dos neurônios, ou seja, os centros dos neurônios e a largura da função gaussiana, conforme mostrado na Eq. (1).

$$\varphi_i(x) = \exp\left(-\frac{\|x_n - u_i\|^2}{2\sigma_i^2}\right) \quad (1)$$

sendo que $x_n = [x_1, x_2, \dots, x_n]$ corresponde ao vetor de entrada, $u_i = [u_1, u_2, \dots, u_i]$ que corresponde ao vetor dos centros e σ corresponde à largura do *i-ésimo* nó, sendo o número de neurônios da camada escondida indicado por n.

Na segunda fase, o treinamento só se inicializa após a finalização da primeira fase, procedimentos de aprendizado supervisionado, pois em função dos valores de erros, utiliza-se a propagação reversa. Nesta fase tem-se os cálculos dos pesos entre a camada escondida e a camada de saída, conforme a Eq. (2).

$$y_n(x) = \sum_{i=1}^n w_i \varphi_i(x) \quad (2)$$

Todo o processo de ajuste dos pesos pode ser verificado na literatura (Haykin, 2001; Silva et al., 2010).

Sistema Adaptativo de Inferência Neuro Fuzzy (ANFIS). A base da lógica de processamento da ANFIS, foi introduzido por Zadeh (1965) e aperfeiçoada por Takagy e Sugeno (1985), e o sistema consiste em utilizar Sistema de Inferência Fuzzy (FIS) juntamente com as técnicas de RNA. Funciona de forma híbrida, combinando características de aprendizagem da RNA com a interpretabilidade da FIS, onde a integração de conhecimento implícito (dados) e explícitos (especialistas) trabalham juntos. Nesse modelo através de interações entre camadas acontece a fuzificação e a defuzificação. Na primeira e segunda camada realiza-se o mapeamento das variáveis por meio da aplicação da função de pertinência, conhecido como fuzificação, onde a regra é agrupar os dados de entrada em subgrupos associando os valores aos subgrupos correspondentes ou com características semelhantes. Na segunda camada representa os antecedentes da regra anterior (fuzzy) aplicando-se uma operação lógica entre as entradas e na terceira a normalização das informações com o grau de ativação da regra anterior em relação as outras. Na quarta camada encontram-se a composição das regras e na quinta o processo de defuzificação com a totalização dos valores. A Fig. 2 mostra a estrutura ANFIS, que é composta por cinco camadas com duas funções de pertinência para cada entrada, descritas como:

- camada 1 e 2: fuzificação: definição dos valores de entrada, da função de pertinência dos nós em relação aos valores de entrada. A saída do nó 2 representa o disparo de um regra e a Eq. 3 é a função de ativação desta camada.

$$w_i = T(\mu_{A_i}(x), \mu_{B_i}(x)), i = 1,2 \quad (3)$$

- camada 3: normalização dos pesos: normalização do disparo realizado anteriormente, dado pela Eq. 4.

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1,2 \quad (4)$$

- camada 4 e 5: defuzificação: cálculo do produto entre a saída da camada 3 e os valores de entrada, ponderados pelos vetores gerados no treinamento (p, q e r), calculada pela Eq. 5.

$$\bar{w}_i z_i = \bar{w}_i (p_i x + q_i y + r_i), i = 1,2 \quad (5)$$

E ao final da camada 5 se realiza o somatório das saídas através da Eq. 6

$$\sum \bar{w}_i z_i = \frac{\sum w_i z_i}{\sum w_i}, i = 1,2 \quad (6)$$

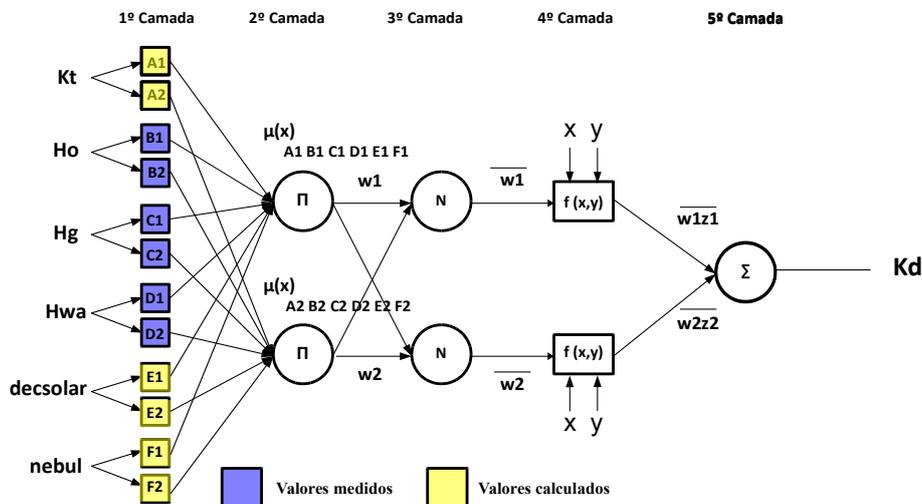


Figura 2: Arquitetura da ANFIS. Adaptado de Jang, 1993.

Todo o processo de arquitetura e ajuste dos pesos pode ser verificado na literatura (Jang, 1993).

Modelo Estatístico. Modelos para estimar a radiação difusa são expressos em fração da radiação com apenas uma variável de referência, a fração transmitida da irradiação global (K_t), pela facilidade de medição das variáveis envolvidas e disponibilidade nas estações meteorológicas. O modelo estatístico proposto (ME) baseou-se na proposta de Liu e Jordan (1960) com apenas uma variável de entrada e na partição horária com equação de estimativa obtida por regressão com dados da localidade medida.

O ME utiliza a relação entre K_d e K_t , sendo $K_d = H_d/H_g$, a fração difusa da irradiação global, onde H_d é a irradiação difusa e H_g é a irradiação global e $K_t = H_g/H_0$, a fração transmitida da irradiação global, conhecida com índice de claridade ou transmissividade atmosférica, onde H_g é a irradiação global e H_0 é a irradiação no topo da atmosfera. Na obtenção do modelo estatístico foram utilizadas as medidas realizadas de H_d e H_g de toda a base de dados com exceção dos valores para validação que foram retirados do conjunto total de informações e dos valores calculados de H_0 .

3. MATERIAL E MÉTODO

3.1 Localização, Instrumentação e Medidas

Os dados para realização deste trabalho foram coletados na Estação de Radiometria Solar do Departamento de Engenharia Rural da Faculdade de Ciências Agrônomicas da UNESP, campus Botucatu, São Paulo (latitude 22°50'48,14" S, longitude 48°25'53,52" W e altitude de 786 m) que pela classificação de Koppen, o clima é do tipo Cwa, temperado úmido, com inverno frio e seco (junho-agosto) e verão quente e úmido (dezembro-fevereiro) e elevada precipitação (Escobedo et al. 2011; Teramoto e Escobedo, 2012). A estação foi criada em 1967 e realiza medidas meteorológicas no mesmo local desde de 1969 sendo que os dados coletados automaticamente se iniciaram somente em 1997 e transmitidos em tempo real a partir de 2014. Os instrumentos utilizados compreendem um piranômetro Eppley – PSP para medidas da radiação global, um pireliômetro Eppley-NIP acoplado a um rastreador solar do tipo ST-3 para medidas da radiação direta, um piranômetro Eppley-PSP com anel de sombreamento para medidas da radiação difusa, calculada a partir do método proposto por Melo e Escobedo (1994) e um pirgeômetro, modelo CG1 da Kipp e Zonen para medidas de radiação de ondas longas. Para medidas da temperatura foi utilizado um termômetro de máximo e de mínimo e a umidade relativa do ar foi medida pelo higrógrafo, instalados dentro de um abrigo meteorológico, sendo todos os equipamentos aferidos pelo método comparativo. Os dados foram coletados entre Janeiro/2000 a Dezembro/2006 e compreendem medidas instantâneas armazenadas a cada cinco segundos, de radiação global, direta, difusa, de ondas longas, temperatura e umidade, sendo utilizadas na pesquisa a média dos últimos cinco minutos das medidas processadas, além dos valores calculados a partir das medidas realizadas, como a nebulosidade e a declinação solar. Os dados coletados foram analisados e valores inconsistentes e discrepantes foram retirados utilizando técnicas praticadas na literatura, como limitação das medidas de radiação global, direta e difusa, como também do ângulo zenital (De Miguel et al., 2001; Lopez et al., 2004; Younes et al., 2005; Ruiz-Arias et al., 2010; Ihya et al., 2015; Pashiardis et al., 2016).

3.2 Treinamento das redes RBF e ANFIS

Treinamento das rede. Para o treinamento das redes RBF e ANFIS foram propostos seis (06) combinações para as variáveis de entrada conforme apresentado na Tab.1, pertencentes ao conjunto de dados medidos e ou calculados descritos na seção 3.1. Estas variáveis foram definidas como entrada para as redes considerando primeiramente as disponíveis e as mais relevantes em função da variável dependente, a irradiação difusa.

Tabela 1: Combinações de entrada para treinamento e validação das redes RBF e ANFIS.

	Modelo		Variáveis de:							
	RBF	ANFIS	Entrada						Saída	
1	RBF 1	ANFIS 1	Kt							K_d
2	RBF 2	ANFIS 2	Kt	Ho						K_d
3	RBF 3	ANFIS 3	Kt	Ho	Hg					K_d
4	RBF 4	ANFIS 4	Kt	Ho	Hg	Hwa				K_d
5	RBF 5	ANFIS 5	Kt	Ho	Hg	Hwa	decsol			K_d
6	RBF 6	ANFIS 6	Kt	Ho	Hg	Hwa	decsol	nebul		K_d

 calculado
 medido

Os valores medidos consistem em valores da irradiação global (H_g) e irradiação de ondas longas atmosféricas (H_{wa}), e dos calculados, a fração transmitida da irradiação global (K_t), da irradiação no topo (H_0), da declinação solar (decsolar) e nebulosidade (nebul). A irradiação solar no topo da atmosfera (H_0) foi calculada de acordo com equações descritas em Iqbal (1983), e utilizadas em vários trabalhos (Souza et al., 2005; Duffie e Beckman, 2013; e Santos et al., 2016), assim como nos cálculos da declinação solar. Para calcular a nebulosidade horária, foi necessário integrar as medidas da irradiância direta (I_b) armazenadas a cada cinco minutos, no período analisado e a partir destas medidas, calcular a

proporcionalidade das medidas horária, da integração da irradiância dos valores abaixo de 120 W/m². Conforme a WMO (2008), em seu guia de instrumentos meteorológicos e métodos de observação, este valor foi adotado com limiar para o brilho do sol, ou seja, para valores inferiores a esse, considera-se sem insolação ou “nebuloso”.

No treinamento das redes RBF e ANFIS foram desenvolvidos dois *scripts* para treinamento e dois para validação da estimativa e os dados processados no *software* MATLAB.

Na parametrização da rede RBF, além dos dados de entrada e de saída, foram definidos também o ajustamento ou *spread* (σ) da função de ativação gaussiana, variando este valor inicialmente no intervalo de 0.001 a 10. Este parâmetro foi definido comparando-se as curvas de dispersão gerada pelos valores estimados da saída da rede RBF e da curva gerada pelos valores estimados do modelo ME com as mesmas variáveis de entrada (Kt e Kd), e que apresentavam o menor erro quadrático médio, que neste caso o valor escolhido foi de 0.01.

Para parametrização da rede ANFIS, além de definir as variáveis de entrada e o ajustamento ou *spread* (σ) da função de ativação, devem ser definidas também a quantidade de funções de pertinência como também o número de interações ou épocas. Neste trabalho definiu-se como função a gaussiana, variando a quantidade de funções do intervalo entre 2 a 4, as interações entre os intervalos de 1 a 200 e o *spread* (σ) no intervalo de 0.001 a 10.

3.3 Base de dados para validação dos modelos estatístico (ME) e redes RBF e ANFIS

A base de dados para validação dos modelos foram retirados da base de treinamento e organizadas em duas bases denominadas de Ano típico (AT) e atípico (AAT). Para definição de cada base foram calculados as médias e o desvio padrão das medidas da irradiação difusa de cada mês da série de todos os anos (Escobedo et al. 2013). Realizada esta fase, foram selecionados para cada mês de toda série entre os anos, os valores mais próximos da média como Ano típico (AT) e os mais afastados da média como Ano atípico (AAT). Esta seleção dos meses que constituem um Ano típico (AT) é similar ao processo de seleção do Ano Meteorológico da Organização Mundial de Meteorologia -WMO (1981). Os meses que constituem um Ano típico (AT) e um Ano atípico (AAT) de toda a série são apresentados na Tab. 2.

Tabela 2: Meses e anos do Ano típico (AT) e Ano atípico (AAT) de janeiro/2000 a dezembro/2006.

		Meses											
		Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Ano	Típico	2003	2000	2001	2001	2000	2001	2005	2001	2006	2006	2005	2001
	Atípico	2002	2003	2000	2004	2004	2002	2004	2000	2005	2005	2001	2005

Em destaque na Tab. 2, observa-se que o ano típico (AT) para o mês de janeiro foi 2003 e para mês de julho foi 2005 e nesses meses o ano atípico (AAT) para o mês de janeiro foi 2002 e para o mês de julho foi o ano de 2004.

3.4 Indicadores estatísticos

Foram utilizados para analisar os resultados encontrados dos modelos, os indicadores rRMSE, raiz quadrada do erro quadrático médio (Root Mean Square Error), o R², coeficiente de determinação e r, coeficiente de correlação (Willmott, 1981; Stone, 1993; Escobedo et al. 2011; Santos et al. 2014). Na avaliação do indicador rRMSE, uma escala classificatória foi proposta com diferentes intervalos como apresentado na Tabela 3 (Jameson et al., 1991; Heinemann et al., 2012).

Tabela 3: Escala de classificação do indicador estatístico

Excelente se:	Bom se:	Aceitável se:	Ruim se:
rRMSE < 10%	10% ≤ rRMSE < 20%	20% ≤ rRMSE < 30%	rRMSE ≥ 30%

4 RESULTADO E DISCUSSÃO

4.1 Modelo estatístico (ME)

Na análise dos resultados do modelo proposto, a curva de dispersão da fração difusa em função do índice de claridade gerada (Fig. 3), é similar aos trabalhos apresentados na literatura assim como a equação de estimativa (Eq. 7) ajustada por regressão polinomial de 4^o ordem com coeficiente de determinação de R² = 0.8 para partição horária.

$$kd = 0.92546 + 1.1164 kt - 4.90289 kt^2 + 1.4679 kt^3 + 1.67489 kt^4 \quad (7)$$

Foram elaboradas equações de outras ordens polinomiais, mas não resultaram em melhoria no coeficiente de determinação no ajuste do modelo. Observa-se através da curva de dispersão da Fig. 2 a), um grande espalhamento dos pontos e em consequência uma grande variabilidade da fração difusa para um mesmo ponto do índice de claridade. Apesar disso, não se pode afirmar que esta situação seja exclusiva deste índice, pois outras variáveis podem contribuir, principalmente na partição horária, que responde mais rápido às mudanças atmosféricas (Dal Pai et al., 2016). Outros elementos podem influenciar nesta variação.

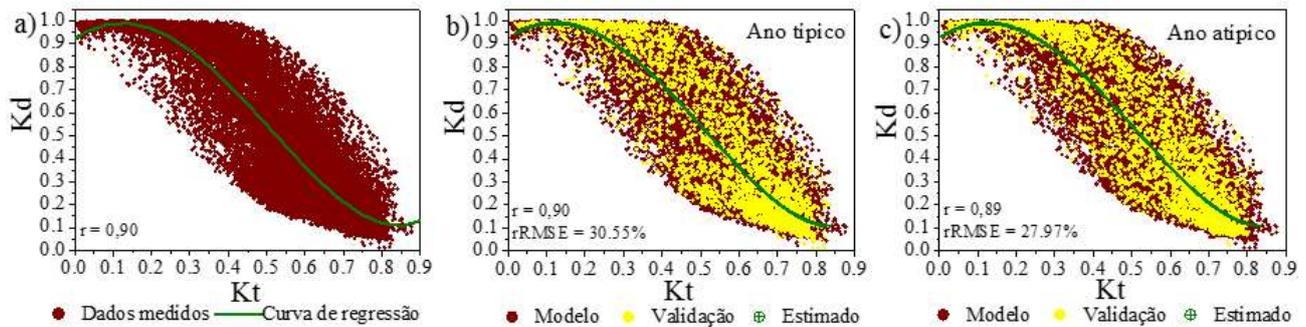


Figura 3: a) Curva de dispersão de Kt em relação a Kd com todos os dados; curva de dispersão de Kd com dados de validação e valor estimado para: b) Ano típico (AT); c) Ano atípico.

Na Tab. 4 são apresentados os indicadores estatísticos resultantes da análise na validação dos modelos ME, e das redes RBF e ANFIS, o rRMSE e o r para Ano típico (AT) e atípico (AAT).

Tabela 4: Indicadores estatísticos do modelo estatístico, das redes RBF 2 e 6 e ANFIS 2 e 6.

Modelo	Variável de entrada	Variável de saída	Topologia e/ou ajustamento	Período de validação - jan/ 2000 a dez/2006			
				Partição: HORÁRIA			
Modelo estatístico				Ano Típico - AT		Ano Atípico - AAT	
				rRMSE	r	rRMSE	r
Modelo empírico							
Modelo - ME	Kt	Kd	Regressão linear	30.55	0.90	27.97	0.89
RBF - Base de Função Radial				rRMSE	r	rRMSE	r
RBF 2	Kt Ho	Kd	0.01	29.63	0.90	26.93	0.90
RBF 6	Kt Ho Hg Hwa decsol nebul	Kd	0.01	26.48	0.92	24.69	0.92
ANFIS - Sistema Adaptativo de Inferência Fuzzy				rRMSE	r	rRMSE	r
ANFIS 2	Kt Ho	Kd		25.13	0.93	22.76	0.93
ANFIS 6	Kt Ho Hg Hwa decsol nebul	Kd		22.63	0.95	21.19	0.94

Na validação do modelo estatístico para AT, uma correlação a princípio satisfatória com $r = 0.90$ (máximo = 1) tendo em vista a partição analisada (horária), mas um erro quadrático ruim ($rRMSE > 30\%$) com um $rRMSE = 30,55\%$, conforme classificação apresentada na Tab. 3. Similar a avaliação anterior, a validação para AAT com coeficiente de correlação $r = 0.89$, o desempenho encontra-se no intervalo aceitável ($20 \leq rRMSE < 30\%$) com um $rRMSE = 27,55\%$.

Para as duas validações existe uma grande variabilidade de Kd em função de Kt que justifica em partes pela partição utilizada mas que outras variáveis astronômicas, geográficas e meteorológicas podem influenciar no espalhamento das medidas analisadas. Apesar de alguns estudos mostrarem esta possibilidade, o índice de claridade (Kt) que participa de todas as combinações, mostra através dos resultados alcançados, ser uma variável importante na estimativa da fração transmitida da irradiação difusa (Paulescu e Blaga, 2016; Furlan et al. 2012)

4.2 Rede de base de função radial (RBF)

Os indicadores estatísticos da rede RBF comparando-os com o modelo tradicional, mostra uma redução de 16% em relação ao erro quadrático dos valores medidos e estimados nas duas bases de validação (AT e AAT) quando todas as variáveis utilizadas estão presentes no processamento dos dados. Pela classificação apresentada na Tab. 3, há um desempenho aceitável ($20 \leq rRMSE < 30\%$) para estimativa de Kd nesta combinação (RBF6). A redução do erro vai diminuindo quando as informações das variáveis astronômicas, geográficas e meteorológicas vão sendo introduzidas nas combinações. A partir da segunda combinação da rede RBF, que utiliza Kt e H_0 como referência, com coeficiente de $r_{AT} = 0.80$ e $r_{AAT} = 0.90$, e um erro de $rRMSE_{AT} = 29.63\%$ e $rRMSE_{AAT} = 26.93$, superior ao valores encontrados no modelo clássico. O melhor resultado é encontrado na combinação quando seis (06) variáveis participam do modelo, com correlação de $r_{AT} = 0.92$ $r_{AAT} = 0.92$ e erro de $rRMSE_{AT} = 26.48\%$ e $rRMSE_{AAT} = 24.65\%$. Na Fig. 4 verifica-se a tendência (4.a. até 4.d.) no comportamento dos valores estimados através da dispersão dos pontos quando as variáveis são inseridas.

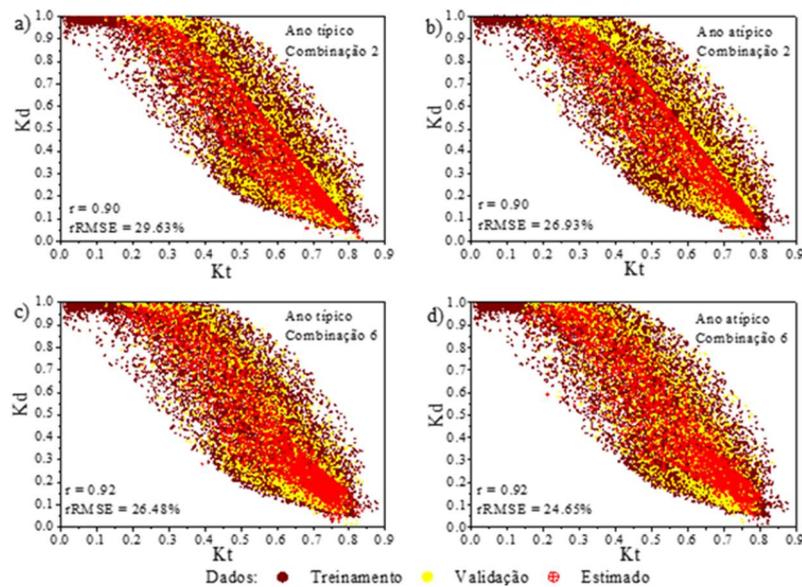


Figura 4: Dispersão da estimativa de K_d calculada pela rede neural do tipo RBF em função de K_t .

4.3 Sistema Adaptativo de Inferência Fuzzy (ANFIS)

Nos indicadores estatísticos da rede ANFIS, a redução foi maior em relação ao modelo tradicional e a rede RBF com uma redução de aproximadamente 34% em relação ao erro quadrático dos valores medidos e estimados nas duas bases de validação (AT e AAT) quando todas as variáveis são utilizadas como entrada da rede. Conforme indicadores (Tab. 3) há um desempenho bom ($10\% \leq rRMSE < 20\%$) para estimativa de K_d nesta combinação (ANFIS6), onde os valores do erro vão diminuindo quando as informações das variáveis astronômicas, geográficas e meteorológicas vão sendo introduzidas através das combinações.

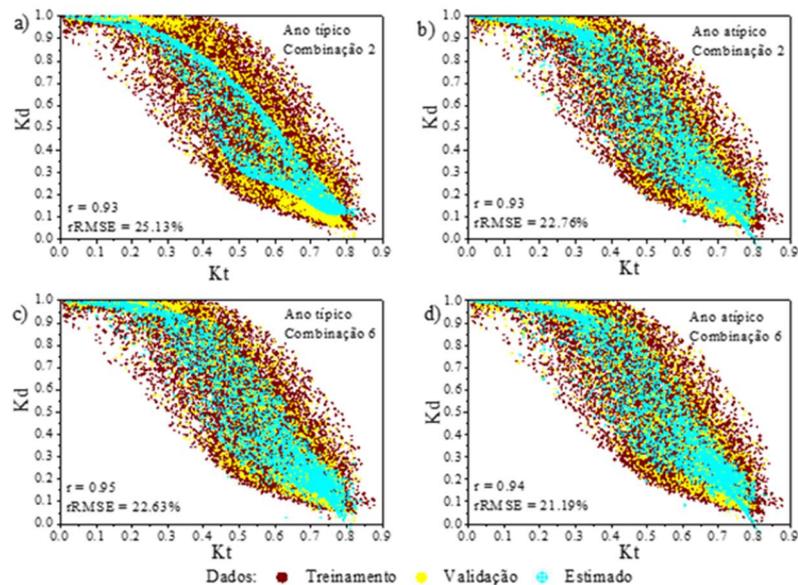


Figura 5: Dispersão da estimativa de K_d calculada pela rede neural do tipo ANFIS em função de K_t .

A partir da segunda combinação da rede RBF, que utiliza K_t e H_0 como referência, com coeficiente de $r_{AT} = 0.93$ e $r_{AAT} = 0.93$, e um erro de $rRMSE_{AT} = 25.13\%$ e $rRMSE_{AAT} = 22.76\%$, superior aos valores encontrados no modelo clássico. O melhor resultado é encontrado na combinação seis com coeficiente de correlação de $r_{AT} = 0.95$ e $r_{AAT} = 0.94$ e erro de $rRMSE_{AT} = 22.63\%$ e $rRMSE_{AAT} = 21.19\%$. Essa redução pode ser verificada na Fig. 5 com a tendência da ANFIS 2 e 6 (5.a. até 5.d.) no comportamento dos valores estimados através da dispersão dos pontos quando as variáveis são inseridas.

2 CONCLUSÃO

O modelo estatístico (ME) proposto nesse trabalho e elaborado a partir de dados coletados da *Estação de Radiometria Solar de Botucatu/SP* e ajustado por regressão polinomial de 4ª ordem, pode ser utilizado para estimativa da

fração difusa da irradiação global (Kd), desde que outros modelos estatísticos com uma ou mais variáveis e com indicadores melhores não sejam elaborados com características geográficas e climatológicas semelhantes. A literatura apresenta alguns modelos com indicadores similares ao apresentado na Tab. 4 (Chandrasekaran e Kumar, (1994); Oliveira et al. (2002); Jacovides et al. (2006); Kuo et al. (2014)) que corrobora com os argumentos de utilizar o modelo ME na estimativa da fração difusa, apesar dos indicadores na validação do modelo (ME) com as duas base de dados, não apresentarem um bom desempenho tanto para a base de validação para Ano típico (AT) e atípico (AAT) conforme classificação apresentada na Tab. 3.

Com a utilização das técnicas de aprendizado de máquinas (RNA/RBF e ANFIS), os indicadores também não apresentaram um ótimo desempenho na estimativa de Kd que conforme os valores encontrados e classificação da Tabela 3 são aceitáveis para estimativa sendo 16% melhor com o uso da técnica RNA/RBF e 34% melhor com o uso da rede ANFIS quando seis (06) variáveis participam do treinamento para as duas bases de validação (RBF6 e ANFIS6). Mesmo com a quantidade de variáveis disponíveis utilizadas como entrada para as redes neste trabalho, as técnicas de aprendizado de máquinas apresentaram melhores resultados, contudo outras combinações e parametrizações devem ser testada a fim de verificar o bom desempenho das técnicas.

Agradecimentos

A UTFPR – Universidade Tecnológica Federal do Paraná, campus Cornélio Procópio e a DAELE – Departamento de Engenharia Elétrica, a Fundação Araucária e a CAPES pelo apoio financeiro parcial ao desenvolvimento deste trabalho.

3 REFERÊNCIAS

- Alam, S.; Kaushik, S. C.; Garg, S. N.. 2009. Assessment of diffuse solar energy under general sky condition using artificial neural network. *Applied Energy* vol. 86, pp. 554–564.
- Azimi, R.; Ghayekhloo, M.; Ghofrani, M.. 2016. A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. *Energy Conversion and Management*, vol. 118, pp.331–344.
- Behrang, M. A.; Assareh, E.; Ghanbarzadeh, A.; Noghrehabadi, A. R.. 2010. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy* vol. 84, pp. 1468–1480.
- Behrang, M. A.; Assareh, E.; Ghanbarzadeh, A.; Noghrehabadi, A. R.. 2010. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy* vol. 84, pp. 1468–1480.
- Benghanem, M.; Mellit, A.. 2010. Radial Basis Function Network-based prediction of global solar radiation data: Application for sizing of a stand-alone photovoltaic system at Al-Madinah. *Energy*, vol. 5, pp. 3751 - 3762.
- Boata, Sta. R.; Gravila, P.. 2012. Functional fuzzy approach for forecasting daily global solar irradiation. *Atmospheric Research* vol. 112, pp. 79 – 88.
- Boland, J.; Scott, L.; Luther, M.. 2001. Modelling the diffuse fraction of global solar radiation on a horizontal surface. *Environmetrics*, vol. 12, pp. 103 –117.
- Bortolini, M.; Gamberi, M.; Graziani, A.; Manzini, R.. Mora, C.. 2013. Multi-location model for the estimation of the horizontal daily diffuse fraction of solar radiation in Europe. *Energy Conversion and Management*, vol. 67, pp. 208–216.
- Chandrasekaran, J.; Kumar, S.. 1994. Hourly diffuse fraction correlation at a tropical location. *Solar Energy* 53, 6, 505 – 510.
- Collares-Pereira, M.; Rabl, A.. 1979. The average distribution on solar radiation – correlations between diffuse and hemispherical and between daily and hourly insolation values. *Solar Energy*, vol. 22, pp. 155 - 164.
- Dal Pai, A.; Escobedo, J. F.; Dal Pai, E.; Oliveira, A. P.; Soares, J. R.; Codato, G.. 2016. MEO shadowring method for measuring diffuse solar irradiance: corrections based on sky cover. *Renewable Energy*, vol. 99, pp.754 - 76.
- De Miguel, A.; Bilbao, J.; Aguiar, R.; Kambezidis, H.; Negro, E.. 2001. Diffuse solar irradiation model evaluation in the north Mediterranean Belt Area. *Solar Energy*, vol. 70, n. 2, pp. 143 – 153.
- Drummond, A. J.. 1956. *Arch. Met. Geoph. B. Serie B.* vol. 7, Issue 3 - 4, pp. 413 – 436.
- Duffie, J. A.; Beckam, W. A.. 2013. *Solar Engineering of Thermal Processes*, Editora Wiley, Canadá.
- Elminir, H. K.; Azzam, Y. A.; Younes, F. I.. 2007. Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models. *Energy*, vol. 32, pp. 1513 – 1523.
- 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*, vol.36, pp.169-178.
- 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*, vol. 36, pp. 169-178.
- Escobedo, J. F.; Santos, C. M.; Dal Pai, A.; Gomes, E. N.. 2013. Estimativa da irradiação solar direta na incidência normal (Kb) em função da transmissividade atmosférica (Kt) em razão da insolação (n/N). *ASADES – Avances en Energia Renovables y Medio Ambiente*, vol. 17, 11.101 – 11.110.

- Furlan, C.; Oliveria, A. P.; Soares, J.; Codato, G.; Escobedo, J. F.. 2012. The role of clouds in improving the regression model for hourly values of diffuse solar radiation. *Applied Energy*, vol. 92, pp. 240 – 254.
- Haykin, S.. 2008. *Redes neurais: princípios e prática*. 2001. 2.ed. Porto Alegre, Bookman.
- 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*, vol. 71, n. 4, pp. 572 - 582.
- Ihya, B.; Mechaqrane, A.; Tadili, R.; Bargach, M. N.. 2015. Prediction of hourly and daily diffuse solar fraction in the city of Fez (Morocco). *Theor Appl Climatol*, vol. 120, pp.737 – 749.
- Iqbal, M. 1983. *An introduction solar radiation*.1983. Livro. Vancouver, Canada. Academic Press.
- Jacovides, C.P.; Tymvios, F. S.; Assimakopoulos, V. D.; Kaltsounides, N. A.. 2006. Comparative study of various correlations in estimating hourly diffuse fraction of global solar radiation. *Renewable Energy*, vol. 31, pp. 2492 – 2504.
- Jameson, 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*, vol. 27, p.337-350.
- Jamil, B.; Siddiqui, A.T.. 2017. Generalized models for estimation of diffuse solar radiation based on clearness index and sunshine duration in India: Applicability under different climatic zones. *Journal of Atmospheric and Solar–Terrestrial Physics* vol. 157–158, pp. 16–34
- Jang, J. S. R..1993. ANFIS : Adap tive-Ne twork-Based Fuzzy Inference System. *Ieee Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3.
- Jiang, Y.. 2008. Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. *Energy Policy*, vol. 36, pp. 3833 – 3837.
- Kambezidis, H. D.; Psiloglou, B. E.; Karagiannis, D.; Dumka, U. C.; Kaskaoutis, D. G.. 2017. Meteorological Radiation Model (MRM v6.1): Improvements in diffuse radiation estimates and a new approach for implementation of cloud products. *Renewable and Sustainable Energy Reviews* vol. 74, pp. 616 – 637.
- Khatib, T.. 2015. A Novel Approach for Solar Radiation Prediction Using Artificial Neural Networks. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 37, Issue 22.
- Khorasanizadeh, H.; Mohammadi. K.; Goudarzi, N.. 2016. Prediction of horizontal diffuse solar radiation using clearness index based empirical models: A case study. *International Journal of Hydrogen Energy*, vol. 41, pp. 21888-21898
- Kisi, O.. 2014. Modeling solar radiation of Mediterranean region in Turkey by using fuzzy genetic approach. *Energy* vol, 64 pp. 429 – 436.
- Kumar, R.; Aggarwal, R.K.; Sharma, J.D.. 2015. Comparison of regression and artificial neural network models for estimation of global solar radiations. *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1294 – 1299.
- Kuo, C.W., Chang, W.C., Chang, K.C.. 2014. Modeling the hourly solar diffuse fraction in Taiwan. *Renewable Energy* vol. 66, pp. 56–61.
- Liu, B. Y. H. e Jordan, R. C.. 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy*, v.3, n.4, pp.1-19.
- Lopez, G.L.; Muneer, T.; Claywell, R.. 2004. Assessment of four shadow band correction models using beam normal irradiance data from the United Kingdom and Israel. *Energy Conversion and Management*, vol. 45, pp.1963 – 1979.
- Melo, J. M. D.; Escobedo, J. F.. 1994. 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.
- Oliveira, A. P.; Escobedo, J. F.; Machado, A. J.; Soares, J.. 2002. Correlation models of diffuse solar-radiation applied to the city of São Paulo, Brazil. *Applied Energy*, vol. 71, pp. 59 – 73.
- Orgill, J. F., Hollands, K. G. T.. 1977. Correlation equation for hourly diffuse radiation on a horizontal surface. *Solar Energy*, vol.19, n.3, pp. 357-359.
- Pashiardis, S.; Kalogirou, S.A.. 2016. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. *Renewable Energy*, vol. 96, pp. 1015 – 1033.
- Paulescu, E.; Blaga, R.. 2016. Regression models for hourly diffuse solar radiation. *Solar Energy*, vol. 125, pp. 111 – 124.
- Quej, V. H.; Almorox, J.; Arnaldo, J. A.; Saito, L.. 2017. 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* vol. 155, vol. 62 -70.
- Robinson, H., Stoch, L.. 1964. Sky radiation and measurements and corrections. *Journal of Applied Meteorology*, v.3, pp.179-81.
- Ruiz-Arias, J.A.; Alsamamra, H.; Tovar-Pescador, J.; Pozo-Vázquez, D.. 2010. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions.. *Energy Conversion and Management*, vol. 51, pp. 881 – 893.
- Santos, C. M.; Escobedo, J. F.; Tadao, E. T.; Silva, S. H. M. G.. 2016. Assessment of ANN and SVM models for estimating normal directirradiation (H_b). *Energy Conversion and Management*, vol.126, pp. 826–836.
- Santos, C. M.; Souza, J. L.; Ferreira Junior, R. A.; Tiba, C.; Melo, R. O.; Lyra, G. B.; Teodoro, I.; Lyra, G.B.; Lemes, M. A. M.. 2014. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. *Energy*, v.71, pp.388-398.
- Shahaboddin Shamshirband, S.; Kasra Mohammadi. K.; Hossein Khorasanizadeh, H.; Yee, P. L.; Malrey Lee, M.; Petković, D.; d,e , Zalnezhad, E.. 2016. Estimating the diffuse solar radiation using a coupled support vector machine–wavelet transform model. *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 428 – 435.
- Silva, I. N.; Spatti, D. H.; Flauzino, R. A.. 2010. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. 1ª edição, São Paulo: Artliber Editora.

- Soares, J.; Oliveira, A. P.; Boznar, M. Z.; Mlakar, P.; Escobedo, J. F.; Machado, A. J., 2004. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique, *Applied Energy*, vol.79, pp. 201-214.
- Souza, J. L.; Nicácio, R. M.; Moura, M. A. L.. 2005. Global solar radiation measurements in Maceió, Brasil. *Renewable Energy*, v. 30, pp. 1203-1220.
- Stone, R.J.. 1993. Improved statistical procedure for the evaluation of solar radiation estimation models. *Solar Energy*, vol. 51, no. 4, pp. 289 – 291.
- Takagi, T., Sugeno, M.. 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* Vol. 15, pp. 116 - 132.
- Teramoto, E. T.; Escobedo, J. F.. 2012. Análise da frequência anual das condições de céu em Botucatu, São Paulo. *Revista Brasileira Engenharia Agrícola Ambiental*. vol.16, n. 9, pp. 985-992.
- Willmott, C. J.. 1982. On the validation of models. *Physical Geography*, vol. 2, pp. 184-194.
- 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.
- WMO – World Meteorological Organization. 2008. Guide to meteorological Instruments and Methods of Observation. WMO-n°8, Seventh Edition, pp.1 - 681, Geneva, Switzerland.
- Younes, S.; Claywell, R.; Muneer, T.. 2005. Quality control of solar radiation data: Present status and proposed new approaches. *Energy*, vol. 30, pp. 1533 – 1549.
- Zadeh, L.A.. 1965. Fuzzy sets. *Inf. Control*. Vol. 8, pp.338–353.

FRACTION ESTIMATION DIFFUSED FROM GLOBAL IRRADIATION WITH MACHINE LEARNING TECHNIQUES

Abstract. Estimates of the diffuse fraction (K_d) were developed as a function of the transmitted fraction of the global irradiation (K_t) in the hourly partition being: Statistical Model (ME); Artificial Neural Networks with Radial Base Function (RBF); and Adaptive Fuzzy Neural Inference System (ANFIS). The statistical model uses as reference only K_t and the techniques a combination of six (06) astronomical, geographic and meteorological variables. The models use a seven-year base (2000-2006) of measurements taken at the Solar Radiometry Station of Botucatu / SP in the hourly partition, being part for training and another for validation divided into AT and AAT. The equation of the statistical model generated by the 4th order polynomial regression presents a coefficient of determination $R^2 = 0.80$ and in the comparison of the measured and estimated values in the validation, a correlation coefficient for a typical year (AT) $r_{AT} = 0.90$ and for the atypical year AAT) of $r_{AAT} = 0.89$, mean square error $rRMSE_{AT} = 30.55\%$ and $rRMSE_{AAT} = 27.97\%$. In the RBF and ANFIS techniques, the models were satisfactory from the second combination, with $r_{AT} = 0.91$ and $r_{AAT} = 0.90$ and $rRMSE_{AT} = 29.63\%$ and $rRMSE_{AAT} = 26.93\%$ and for ANFIS2 a $r_{AT} = 0.93$ and $r_{AAT} = 0.93$ with error $rRMSE_{AT} = 25.13\%$ and $rRMSE_{AAT} = 22.76\%$. For the sixth combination, the RBF6 network had a coefficient of $r_{AT} = 0.92$ and $r_{AAT} = 0.92$ and error of $rRMSE_{AT} = 26.48\%$ and $rRMSE_{AAT} = 24.69\%$ and for ANFIS6 a coefficient $r_{AT} = 0.95$ and $r_{AAT} = 0.94$ and error of $rRMSE_{AT} = 22.63\%$ and $rRMSE_{AAT} = 21.19\%$. The indicators show that the machine learning techniques compared with the statistical model showed a better performance with a reduction of 16% for the RBF6 network and 34% for the ANFIS6 network of the mean square error for two validation bases (AT and AAT) in relation to the ME model.

Key words: Diffuse irradiance, Artificial neural networks of the RBF type; Adaptive Fuzzy Neural Inference System – ANFIS.