

Rede Perceptron Multicamadas (MLP) na estimativa da fração difusa da radiação global

Edson Luis Bassetto (UTFPR) - edbassetto@gmail.com

JANCER FRANK ZANINI DESTRO (UTFPR) - jancer_destro@uol.com.br

Marco Antonio Ferreira Finocchio (UTFPR) - mafinocchio@utfpr.edu.br

Rodrigo Augusto Modesto (UTFPR-CP) - rodrigomodesto@utfpr.edu.br

Adriano de Souza Marques (IFSP) - adriano.marques@ifsp.edu.br

Resumo:

Este trabalho propõe a utilização da Rede Neural Perceptron Multicamadas (RNA/MLP) para estimativa da fração difusa da irradiação global (K_d) na partição diária a partir de um conjunto de combinações de variáveis climatológicas, utilizadas como entrada para treinamento da rede neural. A rede (RNA/MLP) é comparada com modelo estatístico (ME) que utiliza o índice de claridade (K_T) como variável. Os dados utilizados foram obtidos no período de 2000 a 2006 na Estação Meteorológica da Faculdade de Ciências Agrônômicas da Universidade Estadual Paulista, campus Botucatu/SP. Para elaboração do modelo (ME) e treinamento da (RNA/MLP) os dados são reorganizada formando duas bases anuais e independentes para treinamento e validação para ambos os modelos e são denominadas de Ano Típico (AT) e Ano Atípico (AAT). O desempenho é avaliado pelo coeficiente de correlação (r) e o índice de precisão (RMSE). O modelo estatístico (ME) elaborado e ajustado por regressão polinomial de 4ª ordem apresentou um coeficiente de determinação (R^2) de 0,92. O índice de desempenho do modelo estatísticos (ME) na validação com as duas bases (AT e AAT) foi de $RMSE_{AT} = 19,95\%$ e $RMSE_{AAT} = 18,66\%$ e coeficiente de correlação de $r_{AT}=0,95$ e $r_{AAT}=0,95$. Com a (RNA/MLP) os indicadores de desempenho (RMSE) apresentaram uma redução em ambas as bases de validação em aproximadamente 15% a 20% o que mostra que a inserção de variáveis contribui nas estimativas de (K_d). O melhor resultado apresentado pela rede (MLP) é quando todas as variáveis participam do treinamento um $RMSE = 15,56\%$. Estes indicadores mostram que tanto o modelo (ME) quanto a rede (RNA/MLP) suas correlações são similares, contudo a precisão dos valores estimados com a utilização da rede (RNA/MLP) apresentaram resultados mais significativos justificando assim a utilização da técnica como alternativa para as estimativas de (K_d) com as condições de treinamento utilizadas neste estudo.

Palavras-chave: *Radiação Solar; Técnica de Aprendizado de Máquina; Modelo de Estimativa*

Área temática: *Radiação Solar*

Subárea temática: *Recursos Solares e Meteorologia da Radiação Solar*

REDE PERCEPION MULTICAMADAS (MLP) NA ESTIMATIVA DA FRAÇÃO DIFUSA DA RADIAÇÃO GLOBAL

Edson Luis Bassetto – bassetto@utfpr.edu.br

Rodrigo Augusto Modesto – rodrigomadesto@utfpr.edu.br

Jancer Frank Zanini Destro – frank@utfpr.edu.br

Marco Antônio Ferreira Finocchio – mafinocchio@utfpr.edu.br

Universidade Tecnológica Federal do Paraná (UTFPR), Departamento de Elétrica

Adriano Souza Marques – adriano.marques@ifsp.edu.br

Instituto Federal de São Paulo (IFSP), Departamento de Computação

Resumo. Este trabalho propõe a utilização da Rede Neural Perceptron Multicamadas (RNA/MLP) para estimativa da fração difusa da irradiação global (K_d) na partição diária a partir de um conjunto de combinações de variáveis climatológicas, utilizadas como entrada para treinamento da rede neural. A rede (RNA/MLP) é comparada com modelo estatístico (ME) que utiliza o índice de claridade (K_T) como variável. Os dados utilizados foram obtidos no período de 2000 a 2006 na Estação Meteorológica da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista, campus Botucatu/SP. Para elaboração do modelo (ME) e treinamento da (RNA/MLP) os dados são reorganizada formando duas bases anuais e independentes para treinamento e validação para ambos os modelos e são denominadas de Ano Típico (AT) e Ano Atípico (AAT). O desempenho é avaliado pelo coeficiente de correlação (r) e o índice de precisão (RMSE). O modelo estatístico (ME) elaborado e ajustado por regressão polinomial de 4ª ordem apresentou um coeficiente de determinação (R^2) de 0,92. O índice de desempenho do modelo estatísticos (ME) na validação com as duas bases (AT e AAT) foi de $RMSE_{AT} = 19,95\%$ e $RMSE_{AAT} = 18,66\%$ e coeficiente de correlação de $r_{AT}=0,95$ e $r_{AAT}=0,95$. Com a (RNA/MLP) os indicadores de desempenho (RMSE) apresentaram uma redução em ambas as bases de validação em aproximadamente 15% a 20% o que mostra que a inserção de variáveis contribui nas estimativas de (K_d). O melhor resultado apresentado pela rede (MLP) é quando todas as variáveis participam do treinamento um $RMSE = 15,56\%$. Estes indicadores mostram que tanto o modelo (ME) quanto a rede (RNA/MLP) suas correlações são similares, contudo a precisão dos valores estimados com a utilização da rede (RNA/MLP) apresentaram resultados mais significativos justificando assim a utilização da técnica como alternativa para as estimativas de (K_d) com as condições de treinamento utilizadas neste estudo.

Palavras-chave: Radiação Solar; Técnica de Aprendizado de Máquina; Modelo de Estimativa

1. INTRODUÇÃO

A radiação solar é considerada uma das formas de energia mais limpa, sustentável e abundante na superfície terrestre e países com grande incidência dessa energia como o Brasil, são beneficiados com a utilização desta fonte em diversas áreas científicas e tecnológicas (Bakirci, 2015; Teke et al., 2015). No entanto para um aproveitamento eficiente desta radiação é importante quantificar ou estimar este potencial através de medidas (séries temporais ou espaciais) ou por meio de modelos de estimativa.

Dos componentes da radiação solar a irradiância global ($I_g - W/m^2$) ou a irradiação solar global ($H_g, MJ/m^2$) é a energia monitorada com maior frequência nas estações solarimétricas, diferentemente de seus componentes a irradiância direta ($I_b - W/m^2$) ou a irradiação direta ($H_b, MJ/m^2$) importante em projeto de conversão térmica e fotovoltaico (Jamil e Akhtar, 2017) e a irradiância difusa ($I_d - W/m^2$) ou a irradiação difusa ($H_d, MJ/m^2$) importante em estudos de energia passiva e biomassa (Dal Pai et al, 2016; Gueymard, 2008) como também no desenvolvimento de cultivares protegidos. Contudo as medidas da (I_b) que demandam de equipamentos caros são geralmente calculadas pelo método da diferença entre (I_g) e (I_d) pela relação ($I_b = I_g - I_d$). As medidas da (I_d) também não são comuns mas tem sido monitorada e obtidas nos últimos anos com custos menores de equipamentos, operação e manutenção (Dal Pai et al., 2016).

No Brasil as medidas da (H_d) são realizadas em algumas regiões (Marques Filho et al., 2016; Drechmer e Ricieri 2006; Codato et al., 2008; Furlan, et al., 2012; Soares et al., 2004) através de sistemas com baixo custo como na proposta do sistema elaborado por Melo e Escobedo (1994) que está em uma montagem simples através do método com anel de sombreamento. Neste método o anel é fixo a um eixo no plano horizontal e inclinado na latitude local, e o eixo central do anel paralelo a direção polar, diferente de outras montagens (Drummond, 1956 e Robinson e Stoch, 1964; Martinez e Cúnsulo, 2017) que possuem restrições operacionais, de manutenção e custos elevados.

Apesar da importância, medidas ou estimativas da (I_d) ainda são escassas e com poucas informações disponíveis, principalmente no Brasil, país com elevada disponibilidade de recurso solar. (Silva et al. 2008; Oliveira et al., 2002 a e b; Santos et al., 2016a). Diferentemente das medidas (séries temporais ou espaciais) os modelos de estimativa necessitam de informações de outras medidas na superfície terrestre ou atmosférica, como os modelos estatísticos e as técnicas de aprendizado de máquinas. (Jamil e Akhtar, 2017; Khorasanizadeh et al., 2016; Ruiz-Arias et al., 2010).

Vários modelos para estimar (H_d) são estatísticos e relacionam a fração difusa ($K_d = H_d/H_g$) com a transmissividade da irradiação global ou índice de claridade ($K_T = H_g/H_0$) onde (H_0 , MJ/m²) é a irradiação no topo da atmosfera. O primeiro modelo (Liu e Jordan, 1960) correlacionaram estas frações a partir de dados na partição diária. Modelo com uma variável de referência é utilizado na literatura em partições horária e mensal (1977; Collares-Pereira e Rabl, 1979; Chandrasekaran and Kumar, 1994; Oliveira et al., 2002b; Jacovides et al., 2006; Paleuscu e Blaga, 2016) e outros com mais de uma variável (Reindl et al., 1990; Boland et al., 2001; Furlan et al., 2002; Kuo et al., 2014a e b). Uma das desvantagens destes modelos é a restrição ao uso em outros locais de medidas como também outros modelos apresentarem desempenho (RMSE) superiores (Bortolini et al., 2013; Kambezidis et al., 2017) com outras estimativas de radiações.

Outras técnicas recentemente utilizadas nas estimativas são as realizadas por Técnicas de Aprendizado de Máquinas (TAM) do tipo Redes Neurais Artificiais (RNA), Máquinas de Vetores de Suporte (SVM) ou de técnicas híbridas com uso das RNA's e Sistema de Inferência Fuzzy (FIS) entre outras. Estas técnicas tem habilidade em aprender a partir de informações conhecidas e desejadas não conhecendo o comportamento matemático dos dados, com a possibilidade de interpretar falsas informações sem alterar os resultados. Além disto pode estimar resultados a partir de conhecimento previamente de variáveis de grandezas físicas diferentes e ajustar (pesos) os valores das variáveis de entrada de acordo com o valor desejado. Contudo tem a desvantagem na parametrização e ajustes ideais das redes que dependem principalmente da variável e do número de amostras utilizadas.

O uso das TAM para estimativas da radiação solar no Brasil é recente principalmente a radiação difusa (Soares et al., 2004; Marques Filho et al., 2016; Santos et al., 2016b; Silva et al., 2017) sendo necessário a realização de mais estudos e informações disponíveis para a otimização de trabalhos espaciais com radiação no país. Diante da importância de se conhecer a irradiação solar difusa em diversos projetos e estudos de conversão é proposto neste trabalho um estudo não convencional de estimativa através da técnica (RNA/MLP) e prever a fração difusa da irradiação global na partição diária como base em dados astronômicos, meteorológicos e geográficos e comparar com modelo estatístico (ME).

2. REDE NEURAL ARTIFICIAL PERCEPTON MULTICAMADAS (RNA/MLP)

As técnicas de RNA's são modelos computacionais que apresentam uma estrutura formada por um conjunto de elementos chamados neurônios, similar às do cérebro humano, distribuídos paralelamente e composta por camadas de entrada, de saída e camadas ocultas interligadas entre si por conexões.

Modeladas matematicamente através de algoritmos de aprendizado, são capazes de analisar funções complexas e não lineares que adquirem a capacidade de aprendizado e adequação de conhecimento. São utilizadas para o reconhecimento, classificação de padrões, agrupamento, previsão e nos últimos anos na estimativas de variáveis astronômicas. A estrutura básica de uma rede neural é o neurônio artificial ilustrado pela Fig. 1.

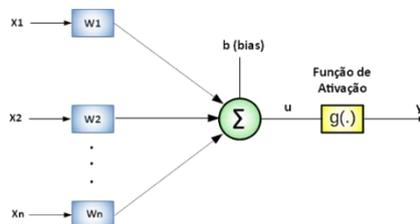


Figura 1 - Modelo do neurônio artificial Fonte: McCullock – Pitts (1943)

É formada por um conjunto de conexões (sinapses) que recebem os valores de entrada (x_1, x_2, \dots, x_i) definidos por pesos sinápticos (w_1, w_2, \dots, w_i) que através de um combinador linear (Σ) executa a soma ponderada de suas entradas, produzindo uma saída chamada de potencial de ativação (u). Na sequência é transformada por uma função de ativação (g), que gera a saída (y) limitada por um intervalo de valores definido pelo limiar de ativação (bias - b) (Haykin, 2001).

Neste trabalho a Rede Neural a ser utilizada é a Perceptron Multicamadas (RNA/MLP) caracterizada pela presença de pelo menos uma camada intermediária de neurônios, localizada entre as camadas de entrada e a camada de saída e são classificadas como redes do tipo *feedforward* (alimentação a frente ou única direção) de camadas múltiplas (Moustris et al., 2008) e o treinamento realizado com o algoritmo Levenberg-Marquadt (LM) (Çelik et al., 2016).

A principal característica de uma (RNA) está na capacidade de aprendizado proporcionado pela arquitetura e no processo de treinamento utilizado. O aprendizado é consequência das mudanças de valores dos pesos sinápticos, conectados aos neurônios entre as camadas com o objetivo de aproximar os valores da saída da rede com os valores desejados (Alam et al. 2009).

Para utilização da (RNA/MLP) é importante a composição da estrutura, a definição do número de neurônios, das camadas ocultas, do algoritmo de treinamento, da função de transferência e do conjunto de dados que podem influenciar nos pesos sinápticos e nos resultados. Com uma arquitetura flexível e pouca complexidade nos algoritmos de treinamento e parametrização são as mais utilizadas na estimativa da irradiação solar. Na estimativa da irradiação global (H_g) (Elizondo et al., 1994; Kalogirou, 2000; Tymvios et al., 2005; Kumar et al., 2015; Silva et al., 2017), com diversas variáveis climatológicas e geográficas como variáveis de entrada. Na irradiação solar direta (H_b) poucos trabalhos são encontrados

na literatura (Alam et. al, 2006; Tomar et. al, 2012; Kaushika et al., 2014; Santos et al., 2016 a e b), assim como para irradiação solar difusa (H_d) (Soares et. al, 2004; Alam et. al, 2009; Ihya et. al, 2014; Boznar et. al, 2017).

3. MATERIAL E MÉTODO

3.1 Localização, Instrumentação e Medidas

A Estação Meteorológica do Lageado está localizada no campus da UNESP – Universidade Estadual Paulista em Botucatu/SP/Brasil (latitude 22,85°S, longitude 48,45°O) e está a 786 m do nível do mar. De acordo com a classificação de Koppen, o clima é do tipo Cwa, com bioma de cerrado e mata atlântica, temperado úmido, com inverno frio e seco (junho-agosto), verão quente e úmido (dezembro-fevereiro) com elevada precipitação (Escobedo et al. 2011; Teramoto e Escobedo, 2012).

Os instrumentos para as medidas instantâneas (irradiância - W/m^2) da irradiância global (I_g - W/m^2) foram obtidas com um piranômetro Eppley-PSP e para medidas da irradiância difusa (I_d - W/m^2) foram utilizado um piranômetro Eppley-PSP com anel de sombreamento (Melo e Escobedo, 1994) com as seguintes especificações: 80 cm de diâmetro, 10 cm de largura e 1mm de espessura. Os valores da (I_d) foram corrigidos através de um fator de correção isotrópico (FC) para compensar a parcela da (I_d) obstruída pelo próprio anel, as quais depende de parâmetros geométricos como largura e raio, geográficos como latitude e os astronômicos como a declinação solar (Oliveira et. al, 2002b).

As medidas obtidas foram armazenadas em um sistema de aquisição de dados automáticos “Datalogger” Campbell Cr23X da Campbell Scientific, operando na frequência de 1Hz. Estes valores são digitalizados a cada 5 segundos, sendo armazenados e transmitido somente as médias realizadas após 5 minutos para um computador através de um módulo de armazenamento SM – 192. Após obtenção das medidas instantâneas (irradiância - W/m^2) os dados foram integralizados em energia (irradiação - MJ/m^2) e separados em séries temporais horárias e diárias (Chaves e Escobedo, 2000). Todos os equipamentos utilizado para as medidas foram aferidos pelo método comparativo a cada dois anos, através de instrumentos adicionais específicos para esta finalidade.

Os dados foram submetidos a uma análise de consistência para determinação dos valores a serem removidos com aplicação de filtros na limitação do ângulo solar (ω_s), da (H_g) em função de H_0 , de (K_T) e (K_d), resultando em uma redução de aproximadamente 7% de todos os dados obtidos (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 Modelo estatístico (ME) e Rede Neural Artificial (RNA/MLP)

Na elaboração do modelo estatístico os valores da fração difusa (K_d) e da transmissividade atmosférica (K_T) calculados a partir de dados experimentais medidos foram correlacionados (Liu e Jordan, 1960) e suas equações de estimativa foram ajustadas por regressão polinomial entre as frações com uma parte maior dos dados e validadas com o restante do conjunto da mesma base. Para treinamento da (RNA/MLP) foram desenvolvidos um *script* 's para treinamento da rede e outro para validação em software específicos para processamento das informações.

Para elaboração do modelo (ME) e treinamento da (RNA/MLP) os dados foram reorganizadas retirando-se uma parte para elaboração e outra para validação do modelo conforme apresentado na Tab. 1.

Tabela 1 - Dados para elaboração, treinamento e validação do ME e da rede MLP

Total de dados (2000 – 2006) 2.348 (100%)	Total de dados para:	
	1. Elaboração do modelo ME	
	1.703 – 72,53%	
	2. Treinamento da rede MLP	
	Base AT	Base AAT
	2.024 – 86,20%	2.027 - 86,33%
	3. Validação do modelo ME e da rede MLP	
	Base AT	Base AAT
324 – 13,80%	321 – 13,67%	

Para validação do modelo, os dados foram reorganizados em duas bases anuais e independentes dos dados para elaboração do modelo denominadas de Ano Típico (AT) e Ano Atípico (AAT) para irradiação difusa (H_d). Esta reorganização foi realizada através de análise estatística da média do valor de (H_d) de cada mês de toda a série assim com o intervalo de variação do desvio padrão. Na sequência comparou-se para cada mês de cada ano o valor de (H_d), a média horária entre os anos, considerando o intervalo de variação do desvio padrão da média total com o valor de (H_d), da média de cada ano. Após análise selecionou-se para cada mês da série (janeiro/2000 - dezembro/2006), os valores próximos a média (acima ou abaixo), como Ano Típico (AT) e mais afastados (acima ou abaixo), como Ano Atípico (AAT).

Diferentemente do modelo ME, que utiliza um conjunto de dados com as variáveis (K_T) e (K_d) para modelagem, no treinamento da técnica (RNA/MLP) são definidos uma sequência de combinações (arranjos) com variáveis climatológicas determinadas a partir dos dados medidos e ou calculados, disponíveis como entrada para as redes de acordo com Tab. 2.

Tabela 2 - Combinações e variáveis de entrada

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

Medido /calculado
 Medido
 Calculado

As variáveis da Tab. 2 consistem em valores de (H_g), irradiação de ondas longas atmosféricas (H_{wa} - MJ/m²), umidade (*umid* - %), temperatura do ar (*tar* - °C) e calculados como (H_0), (K_T), (K_d), declinação solar (*decsol* - °), e nebulosidade (*nebul* - %) calculada em função da proporcionalidade das medidas de (I_b) 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 esses, os mesmos não foram utilizados.

Na implementação da técnica (RNA/MLP) os conjuntos de dados foram submetidos para execução em software que pode ser descrito como uma ferramenta computacional com aplicação em diversas áreas. Permite também o desenvolvimento e análise de sistemas matemático, estatísticos, de otimização e controle com uma linguagem de programação dinâmica e uma ampla disponibilidade de funções para serem implementadas para as técnicas de aprendizado de máquinas. O treinamento da rede (RNA/MLP) foi realizado com o algoritmo Levenberg-Marquadt (*trainlm*) e implementado pela função *feedforwardnet*. Foi parametrizado pelo número de camadas ocultas, do número de neurônios de cada camada, da quantidade máxima de épocas de treinamento e das funções de ativação. A definição do número de camadas foi realizado em duas fases, sendo a primeira com uma camada escondida (n1) e a segunda inserindo-se mais uma camada (n2). O número de neurônios iniciais para cada camada escondida foram definidas em partes atendendo os valores mínimos proposto por Fletcher e Gloss (1993) e apresentados na Tab. 3.

Tabela 3 - Valores para parametrização da rede MLP

Topologias	1	2	3	4	5	6
Nº de neurônios	5	10	20	40	80	160

O número máximo de épocas foi fixado em mil (1.000) interações e a função de ativação nas camadas ocultas a tangente Hiperbólica (*tansig*) definida conforme as características dos dados de entrada e linear na camada de saída. A parametrização da primeira fase foi idêntica à segunda fase. Nas duas fases é realizado dez (10) treinamentos para cada combinação da Tab. 2, testando todas as topologias da Tab. 3. O critério do treinamento é a validação cruzada (70% treinamento, 15% para validação e 15% para teste). Em cada interações (épocas) os erros (MSE – Erro quadrático médio) entre os valores estimados e medidos da validação são avaliados e se forem diminuindo significa que a rede está aprendendo até que este erro comece a aumentar durante uma sequência de seis (6) interações (épocas) sucessivas. Caso este erro da validação no treinamento aumentar enquanto do treinamento continuar caindo é provável que a rede esteja com *overfitting*, condição que indica a parada do treinamento. Este critério minimiza o risco de *underfitting* da rede. Sendo definido a melhor rede em cada treinamento (menor MSE do treinamento) todas as combinações são validadas com os dados reorganizados (AT e AAT) em cada fase e escolhida a fase e a combinação que apresentar o melhor desempenho.

3.4 Indicadores estatísticos

Para analisar os resultados dos indicadores de treinamento e validação das redes foram utilizado o (R^2), coeficiente de determinação, que permite verificar o quanto a variação da variável dependente é explicada pela variação da variável independente através da equação do modelo, (r) coeficiente de correlação que permite verificar o quanto o valor medido se aproxima do valor estimado, ou seja, a correlação existente entre as variáveis, (MSE) erro quadrático médio (Mean Square Error) que indica o erro médio entre os valores estimados e medidos, sendo os melhores resultados aquele com desvios menores e mais próximos dos valores desejados e o RMSE, raiz quadrada do erro quadrático médio (Root Mean Square Error) que permite verificar o valor do erro produzido pelo modelo, sendo que valores menores, indicam um

desempenho satisfatório para os modelos gerados, testados ou treinados e validados, ou seja, mostra o desempenho do modelo quanto ao seu espalhamento (Stone, 1993; Escobedo et al. 2011).

Os indicadores são obtidos pelas seguintes relações:

$$R^2 = \left(\frac{\sum_{i=1}^N (Mi - \bar{M}i)^2 * \sum_{i=1}^N (Ei - \bar{E}i)^2}{\sqrt{\sum_{i=1}^N (Mi - \bar{M}i)^2} * \sqrt{\sum_{i=1}^N (Ei - \bar{E}i)^2}} \right)^2$$

$$r = \left(\frac{\sum_{i=1}^N (Mi - \bar{M}i) * \sum_{i=1}^N (Ei - \bar{E}i)}{\sqrt{\sum_{i=1}^N (Mi - \bar{M}i)^2} * \sqrt{\sum_{i=1}^N (Ei - \bar{E}i)^2}} \right)$$

$$MSE = \left[\frac{\sum_{i=1}^N (Ei - Mi)^2}{N} \right]$$

$$RMSE = 100 * \frac{\left[\frac{\sum_{i=1}^N (Ei - Mi)^2}{N} \right]^{1/2}}{\bar{M}}$$

onde Ei são os valores estimados, $\bar{E}i$, é a média dos valores estimados, Mi são os valores medidos, $\bar{M}i$ é a média dos valores medidos, N é o número de observações.

4. RESULTADOS E DISCUSSÃO

4.1 Modelo estatístico (ME)

O modelo estatístico elaborado da correlação de (K_d) em função de (K_T) da Fig. 2 (a) é similar as curvas encontradas na literatura, assim como as equações de estimativa Eq. (1) 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 elaboradas, mas apresentando coeficiente de determinação menores.

$$K_d = 0,981 + 0,0894 * K_T - 0,462 * K_T^2 - 5,881 * K_T^3 + 5,585 * K_T^4 \dots\dots\dots(1)$$

São apresentados na Fig. 2 (b) e (c) as curvas de dispersão geradas na validação do modelo ME para Ano típico (AT) e atípico (AAT), com os valores dos dados medidos (marrom), dos dados de validação (amarelo) e da estimativa (vermelho).

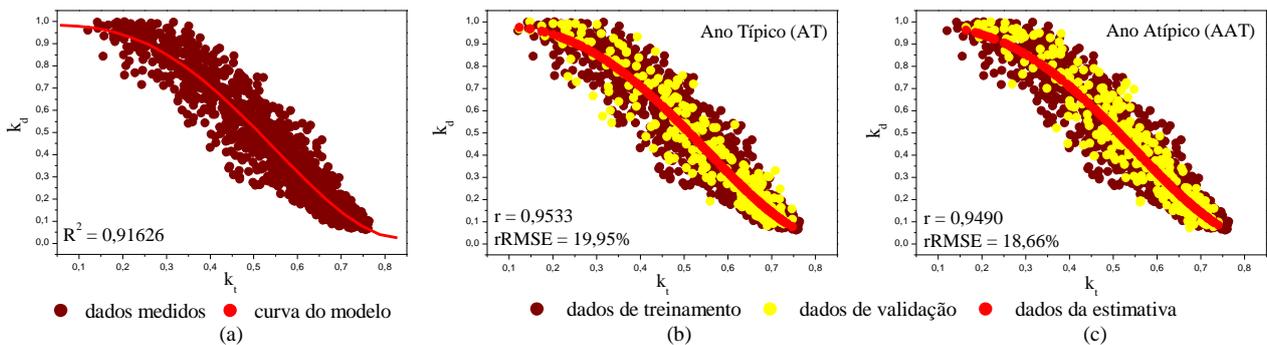


Figura 2 - a) Curva de dispersão ME e validação para: b) AT; c) ATT

Os indicadores estatísticos (RMSE) e o (r) resultantes da análise na validação dos modelos (ME) são para o modelo (ME) na base AT 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 satisfatório ($r = 0.95$) e um erro quadrático bom ($RMSE = 18.66\%$), mostrando que ambos os modelos podem ser utilizados na estimativa da fração difusa na partição diária, considerando a variável disponível na maioria das estações meteorológicas.

Diferentemente de uma análise na partição horária que é mais sensível às mudanças atmosféricas e acontece ao longo do dia, na partição diária essas mudanças são dispersas, fazendo com que essa variação seja menor.

Na Tab. 4 é apresentada a comparação do resultado encontrado do (r) e de (RMSE) da literatura com o melhor resultado do modelo estatístico (ME) para partição diária.

Tabela 4 - 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

Os resultados mostram valores similares ou até superiores nos indicadores de correlação entre os modelos indicando que podem ser utilizados na estimativa da fração difusa, tendo em vista que a elaboração dos modelos seguem as mesmas características construtivas.

4.2 Rede Perceptron Multicamadas

Os indicadores das combinações da técnica (RNA/MLP) é apresentado na Tab. 5 com os melhores resultados experimentados na parametrização e seus indicadores estatísticos após treinamento e validação realizados com a rede para as duas bases de validação.

Tabela 5 - Indicadores e configurações das redes MLP

Configuração e parametrização das Técnicas de Aprendizado de Máquinas - Partição diária																
TÉCNICA MLP																
Para Ano típico (AT)				Indicadores estatísticos					Para Ano típico (AAT)			Indicadores estatísticos				
	Função de ativação	Camada:1/2	Épocas	Treinamento		Validação			Função de ativação	Camada:1/2	Épocas	Treinamento		Validação		
				MSE	RMSE (%)	MSE	RMSE (%)	r				MSE	RMSE (%)	MSE	RMSE (%)	r
Combinação 1	tansig	n1-5/n2-10	19	0.0058	19.29	0.0066	19.98	0.95	tansig	n1-5/n2-40	5	0.0057	19.56	0.0067	18.30	0.95
Combinação 2	tansig	n1-5/n2-40	4	0.0054	18.68	0.0066	19.87	0.95	tansig	n1-5/n2-10	7	0.0054	18.99	0.0065	18.01	0.95
Combinação 3	tansig	n1-10/n2-40	4	0.0054	18.71	0.0065	19.77	0.95	tansig	n1-5/n2-40	4	0.0054	19.03	0.0064	17.92	0.95
Combinação 4	tansig	n1-20/n2-10	4	0.0054	18.66	0.0064	19.66	0.95	tansig	n1-5/n2-10	8	0.0054	18.91	0.0063	17.79	0.95
Combinação 5	tansig	n1-5/n2-20	7	0.0054	18.60	0.0063	19.44	0.96	tansig	n1-20/n2-5	5	0.0053	18.78	0.0067	18.28	0.95
Combinação 6	tansig	n1-5/n2-20	43	0.0043	16.60	0.0052	17.70	0.96	tansig	n1-10/n2-10	10	0.0045	17.25	0.0055	16.59	0.96
Combinação 7	tansig	n1-5/n2-20	10	0.0045	17.12	0.0050	17.34	0.96	tansig	n1-10/n2-10	17	0.0044	17.02	0.0056	16.65	0.96
Combinação 8	tansig	n1-40/n2-10	7	0.0039	15.91	0.0051	17.48	0.96	tansig	n1-10/n2-5	7	0.0044	17.10	0.0048	15.56	0.96

Observa-se nos indicadores de (K_d), que o índice de correlação (r) entre os valores medidos e estimados nas duas base de validação com a rede RNA/MLP variou entre 0,95 a 0,96 mostrando a princípio uma boa correlação com a inserção de variáveis.

No treinamento da rede RNA/MLP o erro quadrático médio (MSE) encontrado foi similar em todas as combinações, sendo que isto pode estar relacionado ao algoritmo de treinamento da técnica. Com a rede (RNA/MLP) o algoritmo do tipo LM proporcionou uma melhor ajustamento dos pesos.

Após realizado todos os treinamentos com a topologia indicada da Tab. 3, são apresentados na Tab. 5 as melhores configurações e parametrização para cada combinação com sua função de ativação, o número de neurônios da primeira e segunda camada, o número de época do melhor desempenho no treinamento e os indicadores estatísticos do treinamento e da validação dos valores estimados em função dos valores medidos. Os índices de desempenho (RMSE) nas duas bases de validação para técnica (RNA/MLP) variou entre os valores 15,56% a 19,98% entre todas as combinações. Observa-se que a (RNA/MLP) apresentou os melhores resultados de validação na base AT quando sete variáveis de entrada (Combinação 7) participaram do treinamento, e quando foram utilizadas duas (2) camadas escondidas. Porém após os dez (10) treinamentos realizados a melhor combinação não foi a que apresentou o menor MSE de treinamento.

Analisando os indicadores da Tab. 5 em relação aos indicadores encontrados com o modelo (ME), a (RNA/MLP) apresentou um desempenho melhor, com uma redução de aproximadamente 20% para base AAT e de 15% para base AT. Na (RNA/MLP), a inserção das variáveis no treinamento contribuí para um melhor desempenho conforme pode ser observado na diminuição do (MSE) de treinamento.

Para um melhor entendimento de exatidão entre os valores medidos e estimados, a Fig. 3 (a) e (b) mostra a dispersão apresentada com a (RNA/MLP) e suas combinações para estimativa da fração difusa nas duas bases AT e AAT.

Importante ressaltar que os resultados encontrados refere-se às combinações proposta e que a alteração dos parâmetros de entrada e da quantidade podem influenciar na configuração das parametrizações e dos indicadores de desempenho das técnicas como no algoritmo de treinamento a ser utilizado onde parâmetros como taxa para aprendizagem e do termo momentum que visa acelerar a convergência ao menor erro entre o valor medido e estimado.

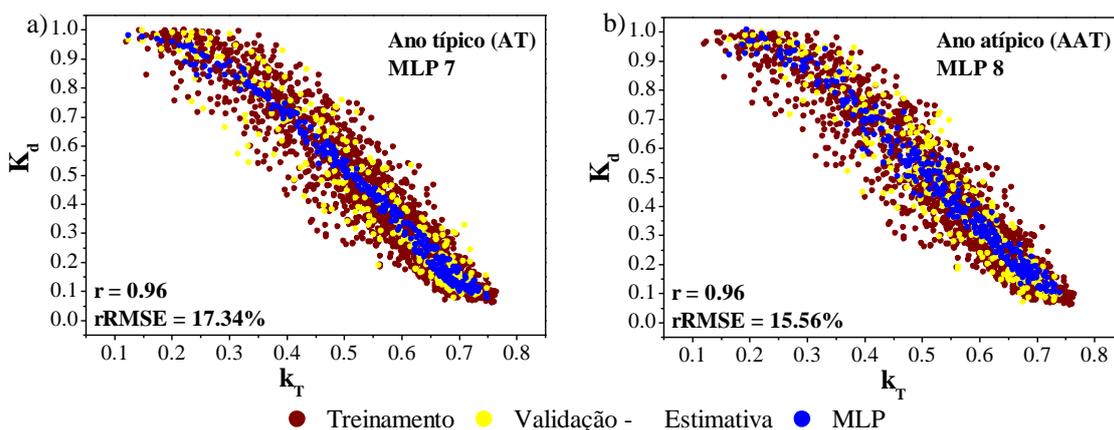


Figura 3 - Curva de dispersão de $K_T \times K_d$ (AT e AAT)

São apresentados na Tab. 6 alguns indicadores do coeficiente de correlação (r) e/ou do valor da raiz quadrado do erro quadrático médio (RMSE) de alguns trabalhos da literatura que utiliza a (RNA/MLP) e o melhor resultado da rede RNA/MLP para partição diária encontrado neste trabalho. Deve-se ressaltar que mesmo com as variáveis de entrada para treinamento das técnicas não serem as mesmas, existem uma forte correlação entre elas.

Tabela 6 – Índices estatísticos - outras localidade e TAM para Boutcatu

Autores	Localidade	Técnicas	Variáveis de entrada	r	RMSE (%)
Elminir et. al (2007)	Cairo/Egito	Perceptron	Radiação global e extraterrestre e brilho solar.		2,52 – 5,96
Xue (2017)	Chengdu, China	Perceptron	Mês do ano, duração da luz do sol, temperatura média, precipitação, velocidade do vento, umidade relativa e radiação solar global diária	0,93	
Ihya et. al (2014)	Marrocos	Perceptron	Transmissividade da irradiação global e fotoperíodo	0,96	18,73
Jiang (2008)	Zhengzhou, China	Perceptron	Índice de claridade	0,96	
Modelo RNA/MLP (AT)	Botucatu/SP	MLP	Transmissividade irradiação global, Radiação no topo da atmosfera, Global e de Ondas longas, declinação solar, nebulosidade, umidade e temperatura.	0,96	17,34
Modelo RNA/MLP (AAT)	Botucatu/SP	MLP	Transmissividade irradiação global, Radiação no topo da atmosfera, Global e de Ondas longas, declinação solar, nebulosidade, umidade e temperatura.	0,96	15,56

Os resultados encontrados são para as combinações e variáveis da Tab. 2 sendo que outras combinações podem ser testadas verificando o comportamento e desempenho das técnicas em seus treinamentos. Com estas mudanças também são necessários outras parametrizações como na escolha do número de camadas, de neurônios, da função de ativação e ajustamento. Com estas mudanças é possível avaliar se a inserção de cada variável irá impactar no desempenho da rede com valores mais significativos para estimativa de (K_d).

Com os resultados apresentados, a (RNA/MLP) mostra-se com uma das Técnicas de Aprendizado de Máquinas na estimativa da fração difusa (K_d) com o uso de variáveis climatológicas e meteorológicas, como também superior aos resultados apresentados pelo modelo ME.

5. CONCLUSÃO

Do que foi apresentado e discutido no presente trabalho é conclusivo que a equação polinomial de 4º (Eq. 1) com $R^2 = 0.92$ pode ser utilizada para estimar a fração difusa da irradiação global com valores de validação de $RMSE_{AT} = 19,95\%$ e $RMSE_{AAT} = 18,66\%$ similares aos encontrados na literatura conforme Tab. 4. A técnica (RNA/MLP) selecionada para estimativa da fração difusa mostraram baixos valores do erro quadrático médio (MSE) nos treinamentos em todas as combinações com inserções de variáveis e as parametrizações propostas para as arquiteturas foram adequadas para redução do erro quadrático médio (MSE) do treinamento e da raiz do erro quadrático médio (RMSE) da validação além de minimizar o tempo e esforço computacional no treinamento da (RNA/MLP).

A validação da RNA/MLP mostraram resultados (valores de RMSE) similares ou superiores aos apresentados na literatura conforme Tab. 6 e a validação da (RNA/MLP) mostraram resultados (valores de RMSE) similares ou superiores

ao (ME) na mesma de condição de entrada (K_T). Além disto a validação da (RNA/MLP) com a inserção das variáveis astronômicas, meteorológicas e geográficas melhoraram o desempenho da técnica na forma sequencial nas arquiteturas propostas e o desempenho (RMSE) da (RNA/MLP) foi superior ao modelo (ME) na maioria das combinações utilizadas.

REFERÊNCIAS

- Alam, S., Kaushik, S. C.; Garg, S. N.. Computation of beam solar radiation at normal incidence using artificial neural network. *Renewable Energy*. v.31, p.1483–1491, 2006.
- Alam, S.; Kaushik, S.C.; Garg, S. N.. Assessment of diffuse solar energy under general sky condition using artificial neural network. *Applied Energy*, vol. 86, p. 554–564, 2009.
- Bakirci, K.. Models for the estimation of diffuse solar radiation for typical cities in Turkey. *Energy*. vol. 82, p. 827 – 838, 2015.
- Boland, J.; Scott, L.; Luther, M.. Modelling the diffuse fraction of global solar radiation on a horizontal surface. *Environmetrics*, vol. 12, p. 103 –117, 2001.
- Bortolini, M.; Gamberi, M.; Graziani, A.; Manzini, R.; Mora, C.. 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.
- Boznar, M. Z.; Bostijan, G.; Oliveira, A. P.; Soares, J.; Mlakar, P.. Spatially transferable regional model for half-hourly values of diffuse solar radiation for general sky conditions based on perceptron artificial neural networks. *Renewable Energy*, vol. 103, p. 794 – 810, 2017.
- Çelik, O.; Teke, A.; Bas, Yildirim, H. B.. The optimized artificial neural network model with Levenberge/Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean. *Journal of Cleaner Production*, vol. 116, pp. 1 -12, 2016.
- Chandrasekaran, J.; Kumar, S.. Hourly diffuse fraction correlation at a tropical location. *Solar Energy*, vol. 53, no 6, p. 505 – 510, 1994.
- Chaves, M., Escobedo, J. F.. A software to process daily solar radiation data, *Renewable Energy*, vol.19 (1), pp. 339 – 344, 2000.
- Codato, G.; Oliveira, A. P.; Soares, J.; Escobedo, J. F.; Gomes, E. N.; Pai, A. D.. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. *Theoretical and Applied Climatology*, vol. 93, pp. 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.; Escobedo, J. F.; Dal Pai, E.; Oliveira, A. P.; Soares, J. R.; Codato, G.. MEO shadowing method for measuring diffuse solar irradiance: corrections based on sky cover. *Renewable Energy*, vol. 99, p.754 - 76, 2016.
- De Miguel, A.; Bilbao, J.; Aguiar, R.; Kambezidis, H.; Negro, E.. Diffuse solar irradiation model evaluation in the north Mediterranean Belt Area. *Solar Energy*, vol. 70, n. 2, p. 143 – 153, 2001.
- Drechmer, P. A. O.; Ricieri, 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 pp. 73 – 77, 2006.
- Drummond, A. J., 1956, *Arch. Met. Geoph. B. Serie B*. vol. 7, Issue 3 - 4, p. 413 – 436, 1956.
- Elizondo, D.; Hoogenboom, G.; McClendon, R. W.. Development of a neural network model to predict daily solar radiation. *Agricultural and Forest Meteorology*, vol. 71, p. 115 - 132, 1994.
- Elminir, H. K.; Azzam, Y. A.; Younes, F. I.. Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models. *Energy*, vol. 32, p. 1513 – 1523, 2007.
- 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.
- Fletcher D.S., Goss, E. . Forecasting with neural network: an application using bankruptcy data. *Information Management*, vol.24, p.159–67, 1993.
- Furlan, C.; Oliveria, A. P.; Soares, J.; Codato, G.; Escobedo, J. F.. The role of clouds in improving the regression model for hourly values of diffuse solar radiation. *Applied Energy*, vol. 92, p. 240 – 254, 2012.
- Gueymard, C.. REST2: high-performance solar radiation model for cloudless sky irradiance, illuminance, and photosynthetically active radiation e validation with a benchmark dataset, *Solar Energy*, vol. 82, pp. 272 – 285, 2008.
- Haykin, S. *Redes neurais: princípios e prática*. 2.ed. Porto Alegre, Bookman, 2001.
- Ihya, B.; Mechaqrane, A.; Tadili, R.; Bargach, M. N.. Prediction of hourly and daily diffuse solar fraction in the city of Fez (Morocco). *Theor Appl Climatol*, vol. 120, pp.737 – 749, 2014.
- Jacovides, C.P.; Tymvios, F. S.; Assimakopoulos, V. D.; Kaltsounides, N. A.. 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.
- Kalogirou, S. Applications of artificial neural-networks for energy systems. *Applied Energy*, vol. 67, p .17 – 35, 2000.
- Kambezidis, H. D.; Psiloglou, B. E.; Karagiannis, D.; Dumka, U. C.; Kaskaoutis, D. G.. 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, p. 616 – 637, 2017.

- Kaushika, N.D.; Tomar, R.K.; Kaushik, S.C.. Artificial neural network model based on interrelationship of direct, diffuse and global solar radiations. *Solar Energy*, vol. 103, pp. 327–342, 2014.
- 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.
- Kumar, R.; Aggarwal, R.K.; Sharma, J.D.. Comparison of regression and artificial neural network models for estimation of global solar radiations. *Renewable and Sustainable Energy Reviews*, vol. 52, p. 1294 – 1299, 2015.
- Kuo, C.W., Chang, W.C., Chang, K.C.. Modeling the hourly solar diffuse fraction in Taiwan. *Renewable Energy* 66, 56–61. 2014a.
- Kuo, C.W., Chang, W.C., Chang, K.C.. Modeling the hourly solar diffuse fraction in Taiwan. *Renewable Energy* 66, 56–61. 2014b.
- Liu, B.H.Y.; Jordan, R. C.. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy*, vol. 4, pp. 1, 1960.
- Lopez, G.L.; Muneer, T.; Claywell, R.. Assessment of four shadow band correction models using beam normal irradiance data from the United Kingdom and Israel. *Energy Conversion and Management*, vol. 45, p.1963 – 1979, 2004.
- Marques Filho, E. P.; Oliveira, A. P.; Willian A. Vita, W. A.; Mesquita, F. L. L.; Codato, G.; Escobedo, J. F.; Cassol, M.; França, J. R. A.. 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.
- Martinez, C.; Cúnsulo, M.. Seguidor automático para bloqueador de radiación solar. *Avances en Energías Renovables y Medio Ambiente*, 2017.
- McCulloch, W., S; Pitts, W.. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, vol. 5, p. 115 – 133, 1943.
- 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.
- Moustris, K.; Paliatsos, A. G.; Bloutsos, A.; Nikolaidis, K.; Koronaki, I.; Kavadias, K.. Use of neural networks for the creation of hourly global and diffuse solar irradiance data at representative locations in Greece. *Renewable Energy*, vol. 33, p 928 – 932, 2008.
- Oliveira, A. P.; Escobedo, J. F.; Machado, A. J.; Soares, J.. 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.
- Pashiardis, S.; Kalogirou, S.A.. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. *Renewable Energy*, vol. 96, p. 1015 – 1033, 2016.
- Reindl, D. T.; Beckman W. A.; Duffie, J. A.. Diffuse fraction correlations. *Solar Energy*, v.45, p.1 – 7, 1990.
- Robinson, H., Stoch, L. Sky radiation and measurements and corrections. *Journal of Applied Meteorology*, v.3, pp.179–81, 1964.
- Ruiz-Arias, J.A.; Alsamamra, H.; Tovar-Pescador, J.; Pozo-Vázquez, D.. 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 JF. 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.; 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*, vol. 126, pp. 826–836, 2016b.
- Shamshirband, S.; Mohammadi, K.; Khorasanizadeh, H.; LipYee, P.; Lee, M.; Petković, D.; Zalnezhad, E.. 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; Escobedo, J. F.; Rossi, T. J.; Santos, C. M.; Silva, S. H. M. G.. 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, pp. 11–23, 2017.
- Soares, J.; Oliveira, A. P.; Boznar, M.Z.; Mlakar, P. Escobedo, J. F.; Machado, A. J.. 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.; Yıldırım, H. B.; Celik, O.. Evaluation and performance comparison of diferent models for the estimation of solar radiation. *Renewable and Sustainable Energy Reviews*, vol. 50, p. 1097 – 1107, 2015.
- 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.
- Tomar, R. K.; Kaushika, N. D., Kaushik, S. C. Artificial neural network based computational model for the prediction of direct solar radiation in Indian zone. *Journal of Renewable and Sustainable Energy*. Vol. 4, p.063 – 146, 2012.
- Tymvios, F. S.; Jacovides, C.P.; Michaelides, S.C.; Scouteli, C.. Comparative study of Ångström’s and artificial neural networks’ methodologies in estimating global solar radiation. *Solar Energy*, vol. 78, p. 752-762, 2005.

WMO – World Meteorological Organization. Guide to meteorological Instruments and Methods of Observation. WMO-n°8, Seventh Edition, p.1 - 681, Geneva, Switzerland, 2008.
Younes, S.; Claywell, R.; Muneer, T.. Quality control of solar radiation data: Present status and proposed new approaches. Energy, vol. 30, p. 1533 – 1549, 2005.

FRACTION ESTIMATION DIFFUSED FROM GLOBAL IRRADIATION WITH MACHINE LEARNING TECHNIQUES

Abstract. This work proposes the use of Perceptron Multilayer Neural Network (RNA / MLP) to estimate the diffuse fraction of global irradiation (K_d) in the daily partition from a set of combinations of climatological variables, used as input for neural network training. The network (RNA / MLP) is compared with a statistical model (ME) that uses only the lightness index (K_T) as a variable. The data used were obtained from 2000 to 2006 at Lageado Meteorological Station of the Faculty of Agronomic Sciences of the Paulista State University, Botucatu campus / SP. For the elaboration of the model (ME) and training of (RNA / MLP) the database is reorganized forming two annual and independent bases for training and validation for both models and are called Typical Year (AT) and Atypical Year (AAT). Performance is assessed by correlation coefficient (r) and precision index (RMSE). The statistical model (SM) elaborated and adjusted by 4th order polynomial regression presented a coefficient of determination (R^2) of 0.92. The performance index of the statistical model (SM) in the validation with both bases (AT and AAT) was $RMSE_{AT} = 19.95\%$ and $RMSE_{AAT} = 18.66\%$ and correlation coefficient of $r_{AT} = 0.95$ and $r_{AAT} = 0.95$. With (RNA / MLP) the performance indicators (RMSE) showed a reduction in both validation bases by approximately 15% to 20% which shows that the insertion of variables contributes to the estimates of (K_d). network (RNA / MLP) is when all variables participate in training an $RMSE = 15.56\%$. These indicators show that both the model (SM) and the network (RNA / MLP) their correlations are similar showing that all variables are well correlated, however the accuracy of the estimated values in relation to the values measured with the use of the net (RNA / MLP) presented more significant results justifying the use of the technique as an alternative to the estimates of (K_d) with the training conditions used in this one. study.

Key words: Solar radiation; Machine Learning Technique; Estimate Model.