

TÉCNICA DE APRENDIZADO DE MÁQUINAS PARA ESTIMAR A FRAÇÃO DIFUSA DA RADIAÇÃO GLOBAL COM BASE EM VARIÁVEIS CLIMATOLÓGICAS PARA UMA ZONA CLIMÁTICA SUBTROPICAL COM BIOMA DE MATA ATLÂNTICA, ÚMIDO MESOTÉRMICO.

Edson Luis Bassetto – bassetto@utfpr.edu.br

Rodrigo Augusto Modesto

Wagner Fontes Godoy

Angelo Feracin Neto

Jancer Frank Zanini Destro

Marco Antonio Ferreira Finocchio

Marcio Mendonça

Universidade Tecnológica Federal do Paraná, Campus Cornélio Procopio

Adriano Souza Marques

Marcos Roberto Ruybal Bica

Instituto Federal de São Paulo, Campus Birigui

Resumo. O trabalho propõe a utilização de uma Técnica de Aprendizado de Máquinas para estimativa da fração difusa (K_a) da irradiação global solar a partir de um conjunto de combinações de variáveis astronômicas e meteorológicas medidas e calculadas que foram extraídos no período de 2020 a 2022 de uma das seis Estações de Pesquisa em Energia Solar do Laboratório de Energia Solar – Labens, da Universidade Tecnológica Federal do Paraná. A técnica de aprendizado de máquina utilizado são Redes Neurais Artificiais (RNA/MLP) e para verificar o desempenho da técnica é elaborado um modelo estatístico (ME) utilizando somente a fração transmitida da irradiação solar global (K_T) como variável de referência. As variáveis de entrada para treinamento das RNA/MLP compreendem valores de nove variáveis astronômicas e meteorológicas medidas e calculadas. O desempenho é avaliado pelo coeficiente de correlação (r) e o índice de precisão (RMSE) e o modelo estatístico (ME) elaborado e ajustado por regressão polinomial de 4ª ordem e apresenta um coeficiente de determinação (R^2) de 0,7919 o que mostra que K_a está correlacionado com K_T estatisticamente. O índice de desempenho do modelo estatísticos (ME) na validação com a base de dados foi de RMSE = 30,7015% o que mostra uma boa aproximação dos valores estimados aos medidos e de correlação de $r=0,8899$ o que mostra a importância das variáveis independente para variável dependente. Pode-se observar nos resultados que com as RNA/MLP os indicadores de desempenho (RMSE) apresentaram uma redução em relação ao ME em aproximadamente 23%, o que mostra que a inserção de variáveis contribui nas estimativas de (K_a) com as condições astronômicas, meteorológicas e geográficas utilizadas.

Palavras-chave: Radiação difusa; Aprendizado de Máquinas; Perceptron Multicamadas.

1. INTRODUÇÃO

A radiação solar é considerada uma forma de energia limpa, sustentável e abundante na superfície terrestre e que tem sido objeto de muitas discussões principalmente em atender uma demanda crescente das matrizes energéticas de muitos países, principalmente aqueles com grande incidência dessa energia como o Brasil, que podem ser beneficiados com o uso desta fonte em diversas áreas científicas e tecnológicas (Bakirci, 2015; Teke *et al.*, 2015). Porém é importante quantificar ou estimar este potencial através de medidas ou por meio de modelos de estimativa para verificar e avaliar a disponibilidade desta energia ao longo do tempo em uma determinada região para um bom aproveitamento.

Das componentes da radiação solar, medidas da irradiância solar global ($I_g - W/m^2$) é a irradiância monitorada com maior frequência nas estações solarimétricas, diferentemente das medidas da irradiância solar direta na horizontal ($I_{bn} - W/m^2$) e da irradiância solar difusa na horizontal ($I_d - W/m^2$) que são importantes em projetos de conversão fotovoltaico e térmico (Jamil e Akhtar, 2017) e importante em estudos de energia passiva e biomassa (Dal Pai *et al.*, 2016; Gueymard, 2008) e também no desenvolvimento de cultivares protegidos (Escobedo *et al.*, 2013). Mas medidas de (I_b) que demandam de equipamentos custosos geralmente são 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 e são realizadas em poucas regiões principalmente em países em desenvolvimento como o Brasil, mas podem ser monitoradas com custos menores de equipamentos como o anel de sombreamento (Dal Pai *et al.*, 2016; Marques Filho *et al.*, 2016; Drechmer e Ricieri 2006). Um destes equipamentos foi elaborado e construído por Melo e Escobedo (1994) e consiste de uma montagem simples onde 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 com o sensor que

transladando sobre a horizontal nesse plano, 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 sua importância, medidas ou estimativas da radiação difusa ainda são escassas e poucas informações são disponíveis, principalmente no Brasil, país com elevada disponibilidade de recurso solar. (Silva *et al.*, 2008; Oliveira *et al.*, 2002; Santos *et al.*, 2016). Diferentemente das medidas em séries temporais ou espaciais, os modelos de estimativa são uma alternativa nos últimos anos para disponibilizar estas medidas, mas necessitam de informações de outras medidas na superfície terrestre como os modelos estatísticos e as técnicas de aprendizado de máquinas. (Jamil e Akhtar, 2017; Khorasanizadeh *et al.*, 2016; Ruiz-Arias *et al.*, 2010).

Na maioria, os modelos para estimar a irradiação solar difusa ($H_d - MJ/m^2$) são estatísticos e para minimizar os efeitos das variações temporais é proposto a utilização de frações que permitem em partes minimizar alguns efeitos astronômicos e geográficos. Os primeiros trabalhos a correlacionaram a fração da irradiação solar difusa ($K_d = H_d/H_g$) onde ($H_g - MJ/m^2$) é a irradiação global, com a fração transmitida da irradiação solar global ou índice de claridade ($K_T = H_g/H_0$) onde ($H_0 - MJ/m^2$) é a irradiação no topo da atmosfera, foram propostos por Liu e Jordan (1960) a partir de uma base de dados destas variáveis na partição diária.

Outros modelos com apenas uma variável de referência são utilizados na literatura em partições horária, diária e mensal (Chandrasekaran and Kumar, 1994; Oliveira *et al.*, 2002; Jacovides *et al.*, 2006; Paleuscu e Blaga, 2016), como também outros com mais de uma variável (Reindl *et al.*, 1990; Ridley *et al.*, 2010; Furlan *et al.*, 2012; Kuo *et al.*, 2014). Além destes outros modelos de estimativas estão sendo aplicados como as Técnicas de Aprendizado de Máquinas (TAM) do tipo Redes Neurais Artificiais (RNA), Máquinas de Vetores de Suporte (SVM), e de técnicas híbridas com uso das RNA's e Sistema de Inferência Fuzzy (FIS) entre outras. Estas técnicas permitem utilizar diferentes parâmetros de grandeza físicas diferentes e através de algoritmos prever medidas a partir de uma base de dados conhecida (Khatib *et al.*, 2012; Paulescu *et al.*, 2016).

Diante destas considerações e da importância de se conhecer as medidas da radiação solar difusa na superfície horizontal pretende-se utilizar a TAM (RNA/MLP) na estimativa da fração da irradiação solar difusa (K_d) na partição horária como base em dados climatológicos obtidos na estação da cidade de Campo Mourão/Pr e ao final compará-la com modelo estatístico (ME) de referência elaborado com a mesma base de dados e com apenas uma variável independente.

2. REDES NEURAIS ARTIFICIAIS (RNA)

As Redes Neurais Artificiais (RNA) são modelos computacionais que apresentam uma estrutura formada por um conjunto de elementos chamados neurônios, distribuídos paralelamente e composta por camadas de entrada, de saída e camadas ocultas interligadas entre si por conexões. Essas redes modeladas matematicamente através de algoritmos de aprendizado são capazes de analisar funções complexas e não lineares que adquirem a capacidade de aprendizado e adequação. São utilizadas para diversos fins científicos, mas nos últimos ano em estimativas de variáveis astronômicas.

A estrutura básica de uma rede neural é o neurônio artificial ilustrado pela Fig. 1. Formada por um conjunto de conexões (*sinapses*) que recebem os valores de entrada (x_1, x_2, \dots, x_n) 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 (f), que gera a saída (y) limitada por um intervalo de valores definido pelo limiar de ativação (*bias* - b) (Haykin, 2001).

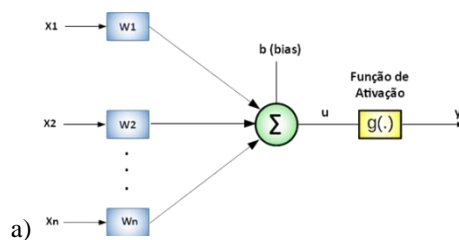


Figura 1 - Modelo do neurônio artificial - Fonte: McCullock – Pitts (1943)

A RNA a ser utilizada é a Perceptron Multicamadas (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 treinada com o algoritmo *Levenberg-Marquadt* (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), 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), assim como para irradiação solar difusa (H_d) (Alam *et al.*, 2009; Ihya *et al.*, 2014; Boznar *et al.*, 2017).

3. MATERIAL E MÉTODOS

3.1 Localização e Clima

A Estação Solarimétrica da UTFPR do campus de Campo Mourão/Paraná faz parte de um projeto de pesquisa desenvolvido pela Universidade Tecnológica Federal do Paraná – UTFPR em parceria com a Companhia Paranaense de Energia Elétrica (COPEL) onde foram instalados uma rede de Estações de Pesquisa em Energia Solar e coordenado pelo Laboratório de Energia Solar – Labens da UTFPR. A rede consiste de Estações Solarimétricas e módulos de avaliação instalados nos campi da universidade em Curitiba, Ponta Grossa, Pato Branco, Medianeira, Campo Mourão e Cornélio Procopio. Neste trabalho a estação onde os dados foram extraídos está localizada em Campo Mourão/Pr (Latitude: 24.046 S, Longitude: -52.3830) e está a 596 m do nível do mar. De acordo com a classificação de Koppen, o clima é do tipo subtropical (Cfa) com bioma de mata atlântica, clima úmido mesotérmico, com verão quente com médias de temperatura superiores a 22° C e o mês mais frio com temperatura inferior a 18° C.

3.2 Obtenção da Base de dados

Os dados medidos foram obtidos simultaneamente da irradiância global na horizontal ($I_g - W/m^2$), da irradiância global inclinada ($I_{g\ incl} - W/m^2$), da irradiância difusa (I_d), da irradiância direta ($I_{bn} - W/m^2$), da irradiância de ondas longas ($I_{wl} - W/m^2$), da irradiância fotossinteticamente ativa ($I_{par} - W/m^2$) e da temperatura (temp) no período de 2020 a 2022.

Nas medidas instantâneas (W/m^2) da irradiância solar global na horizontal e no plano inclinado (I_g e $I_{g\ incl}$) foi utilizado um Piranômetro Modelo CMP 10 da Kipp&Zonen, e para medidas da irradiância solar direta normal (I_{bn}) foi usado um Pireliômetro CHP 1 da Kipp&Zonen acoplado a um rastreador solar do tipo Solys 2 da Kipp&Zonen e para medidas da irradiância solar difusa (I_d) um Piranômetro com anel de sombreamento Modelo CMP 10 da Kipp&Zonen. Para medidas de irradiância de ondas longas atmosféricas (I_{wl}), um pirgeômetro, modelo CGR4 da Kipp&Zonen, e da irradiância Fotossinteticamente Ativa (I_{par}) um sensor PAR Quantum modelo PQS 1 da Kipp&Zonen e para medidas de temperatura do ar (tp) (°C) um Termohigrômetro modelo 41382VC TEMP/RH SENSOR, DEG C da marca Young. Após obtenção das medidas instantâneas (unidade - W/m^2), foram calculadas as médias horárias para utilização neste trabalho.

Todas as medidas obtidas passaram por um controle de qualidade, adotando-se aqui alguns testes de procedimentos como limitação de irradiância global horária (I_g) em função da irradiância no topo da atmosfera (I_0), da fração transmitida da irradiância global (K_T) e da fração difusa da irradiância global (K_d) (Younes *et al.*, 2005; Ruiz-Arias *et al.*, 2010; Ihya *et al.*, 2014; Pashiardis e Kalogirou, 2016). Foram retirados aproximadamente 7,4% do total de dados medidos do período.

3.3 Modelo Estatístico (ME)

Para elaboração do Modelo Estatístico (ME) os valores da fração da difusa da irradiância global ($K_d = I_d/I_g$) e da transmissividade atmosférica ($K_T = I_0/I_g$) foram calculados a partir de dados experimentais medidos nas Estações do LABENS do campus Campo Mourão que foram correlacionados (Liu e Jordan, 1960) e suas equações de estimativa ajustadas por regressão polinomial entre as frações.

3.4 Elaboração e validação do Modelo Estatístico (ME)

Na elaboração do ME e para utilização nos treinamentos das redes neurais os dados foram reorganizados retirando-se uma parte para treinamento e outra para validação do modelo conforme apresentado no item 3.6 sendo os valores utilizados uma média das medidas instantâneas no período de uma hora. Neste trabalho os valores horários não foram integrados.

3.5 Técnicas de aprendizado de máquinas

Para treinamento da rede foram desenvolvidos *script's* para treinamento e validação, com diferentes funções e parametrização sendo sua validação realizada com a base de acordo com cada descrição das variáveis apresentada na Tab. 1. Para modelagem, no treinamento da RNA/MLP são definidos uma sequência de combinações com variáveis climatológicas determinadas a partir dos dados medidos e ou calculados.

Tabela 1: Variáveis selecionadas para as estimativas com as técnicas de Aprendizado de Máquina.

Descrição	Variável	Medida	Calculada	
Fração da Radiação Global na superfície horizontal	K_T		X	V1
Radiação de Topo	I_0		X	V2
Radiação Global	I_g	X		V3
Radiação Direta	I_{bn}	X		V4
Radiação de Ondas Longas	I_{wl}	X		V5
Radiação PAR	I_{par}	X		V6
Declinação Solar	δ		X	V7
Radiação Global em superfície inclinada	I_{gincl}	X		V8
Temperatura Média na superfície	tp	X		V9
Fração da Radiação Difusa na superfície horizontal	K_d		X	V10

Os dados medidos e ou calculados foram disponíveis como entrada para as redes de acordo com Tab. 2, diferentemente do modelo estatístico, que utiliza um conjunto de dados somente com as variáveis K_T e K_d .

Tabela 2: Arranjos das variáveis e treinamento da RNA.

Variáveis de entrada									Variável de saída
V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
K_T	I_0	I_g	I_{bn}	I_{wl}	I_{par}	δ	I_{gincl}	tp °C	K_d
C1									
C2									
C3									
C4									
C5									
C6									
C7									
C8									
C9									

Vários arranjos foram elaborados para o treinamento da rede contudo em função das condições de seleção das variáveis, as mesmas não foram utilizadas sendo que neste estudo optou-se em utilizar variáveis considerando principalmente sua disponibilidade, sendo os valores medidos de I_g , irradiância solar global na horizontal (W/m^2), de I_{bn} , irradiância solar direta (W/m^2), de I_{wl} , irradiância de ondas longas atmosféricas (W/m^2), de I_{par} , irradiância fotossinteticamente ativa (W/m^2), I_{gincl} irradiância solar global inclinada (W/m^2) e tp , temperatura (T , °C) e calculados como I_0 , irradiância no topo da atmosfera (W/m^2), de transmissividade atmosférica ou índice de claridade (K_T) e fração difusa (K_d), e δ , declinação solar (°C).

Na Tab. 3 são apresentados uma parte dos dados e o arranjo do arquivo no formato em CSV, sendo que as variáveis de entrada e saída foram organizadas em arranjos de combinações, separadas em arquivos no formato CSV (Comma-Separated Values) contendo os dados de treinamento.

Tabela 3: Exemplo da estrutura do arquivo CSV.

K_T	I_0	I_g	I_{bn}	I_{wl}	I_{par}	δ	I_{gincl}	tp °C	K_d
0.82	1126.98	925.30	936.99	126.41	1972.87	-6.38	933.65	0.00	0.09
0.78	1249.91	978.30	944.49	130.27	2089.01	-6.38	997.16	25.79	0.09
0.82	1291.83	1059.37	938.29	134.25	2271.17	-6.38	1097.75	27.03	0.13
0.83	1249.91	1034.32	792.16	120.27	2203.24	-6.38	1060.09	28.74	0.25
0.51	1126.98	577.45	355.68	100.18	1232.42	-6.38	589.12	29.02	0.44
0.72	931.44	674.81	583.15	94.26	1424.21	-6.38	687.74	29.55	0.37
0.96	676.60	648.54	808.86	111.86	1367.29	-6.38	663.61	30.53	0.29
0.88	379.83	335.16	674.46	117.63	715.61	-6.38	325.74	30.29	0.30
0.91	61.36	55.89	161.96	115.44	123.67	-6.38	58.34	28.49	0.39
0.32	57.64	18.29	86.20	71.51	38.27	-5.99	11.41	16.54	0.53
0.51	376.42	191.81	556.32	84.64	420.21	-5.99	148.14	18.61	0.26

As combinações dos arranjos foram determinadas de forma a promover uma avaliação da influência das variáveis selecionadas nos resultados das estimativas.

Definido as combinações das variáveis de entrada apresentam-se na sequência a metodologia para treinamento da rede.

3.6 Rede Perceptron Multicamadas (MLP)

A técnica de Rede Neural do Tipo Perceptron foi implementada com um conjunto de dados que foram submetidos para execução no software MATLAB R2014. O treinamento da rede MLP foi realizado com o algoritmo *Levenberg-Marquadt* (*trainlm*) e implementado pela função *feedforwardnet* e parametrizado pelo número de camadas ocultas, do número de neurônios, da quantidade máxima de épocas de treinamento e das funções de ativação. A definição do número de camadas foi realizada com apenas uma camada escondida (n_1). O número de neurônios iniciais para cada camada escondida foi definido em partes atendendo os valores mínimos proposto por Fletcher e Gloss (1993). O número máximo de épocas foi fixado em mil (1.000) interações e a função de ativação Tangente Hiperbólica (*tansig*) definida conforme as características dos dados de entrada. Na parametrização são realizados cinco (5) treinamentos para cada combinação e o critério de parada do treinamento é a validação cruzada (70% treinamento, 15% para validação e 15% para teste).

A condição para determinação do número de neurônios depende de uma série de fatores, como a quantidade de dados disponível, a qualidade dos dados e complexidade do problema (não-linear, descontínuo etc.), mas de forma geral o número de neurônios é determinado de forma empírica com base na capacidade de generalização da rede. Neste estudo identificou-se a quantidade de 25 neurônios e apenas uma camada oculta: Os treinamentos foram realizados em cada combinação, as quais são treinadas consecutivamente. Desse conjunto de treinamento são calculadas as médias do EQM obtidos, de forma a se escolher a rede que apresentou o menor valor médio do erro.

O critério de parada do treinamento da MLP foi a validação cruzada, sendo que no processo de treinamento os dados são separados aleatoriamente em 70% para treino, 15% para validação e 15% para teste. A cada ciclo (época) os erros da validação são avaliados, e enquanto continuar diminuindo significa que a rede continua aprendendo.

Em cada interações (épocas) os erros (EQM – 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 EQM do treinamento) todas as combinações são validadas com os dados em cada fase e escolhida a fase e a combinação que apresentar o melhor desempenho.

A definição da melhor combinação para estimativa de K_a se dá pela escolha do menor resultado do índice de precisão rRMSE da validação cruzada.

3.7 Indicadores de Validação

Para validação do modelo foram utilizados o EQM - erro quadrático médio (*Mean Square Error*) que indica o erro médio entre os valores estimados e medidos, o indicador RMSE, raiz quadrada do erro quadrático médio (*Root Mean Square Error*) que permite verificar o valor do erro produzido pelo modelo, o R^2 , coeficiente de determinação, permite verificar o quanto a variação da variável dependente é explicada pela variação da variável independente através da equação do modelo, e r , coeficiente de correlação que permite verificar o quanto o valor medido explica o valor estimado, ou seja, a correlação existente entre eles, sendo os valores mais próximos de 1 os mais satisfatórios. Para o EQM os melhores resultados são aquele com desvios menores mais próximos dos valores desejados e do RMSE os valores menores, indicam um desempenho satisfatório para os modelos gerados. (Willmott, 1982; Stone, 1993; Escobedo *et al.*, 2011).

Os indicadores são obtidos pelas seguintes relações:

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

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 DISCUSSÕES

Neste item são apresentados os resultados obtidos para estimativa de K_d utilizando o modelo estatístico com uma variável e na sequência o resultado obtido por meio das técnicas de Rede Neural/MLP com variáveis climatológicas de referência e na sequência uma análise de desempenho dos modelos estatístico e da rede.

4.1 Modelo Estatístico

O modelo estatístico elaborado da correlação de K_d em função de K_T foi ajustado por regressão polinomial de 4ª ordem conforme Eq. (1) e com coeficiente de determinação na ordem $R^2 = 0.7919$.

$$K_d = 3.1968 * K_t^4 - 1.1852 * K_t^3 - 3.5395 * K_t^2 + 1.1002 * K_t + 0.8317 \quad (1)$$

Outras equações de ordens polinomiais foram elaboradas, mas apresentando coeficiente de determinação menores. Na Fig. 2 é apresentada a curva de dispersão do modelo estatístico

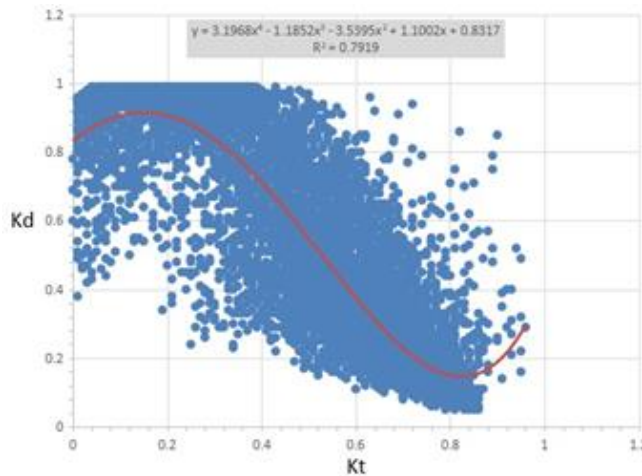


Figura 2: Curva de dispersão do modelo estatístico.

4.2 Rede Neural Artificial MLP

Após o processo de treinamento da rede MLP parametrizada conforme descrito no item 3.6, as redes com os melhores resultados foram organizadas apresentando o erro quadrático médio, o número de épocas e da quantidade de neurônios utilizados como apresentado na Tab. 4.

Tabela 4: Erro quadrático médio, número de épocas, número de neurônios e camada oculta do treinamento da RNA.

	Comb	Variáveis de Entrada										Variável Estimada	EQM	Épocas	Número de Neurônios	Camada Oculta	
Estatístico	ME	K_t											K_d				
RNA / MLP	C1	K_t												0.0196	6	25	1
	C2	K_t	I_0									0.0131		20			
	C3	K_t	I_0	I_g							0.0129	10					
	C4	K_t	I_0	I_g	I_{bn}					0.0014	23						
	C5	K_t	I_0	I_g	I_{bn}	I_{wl}				0.0015	36						
	C6	K_t	I_0	I_g	I_{bn}	I_{wl}	I_{par}			0.0014	32						
	C7	K_t	I_0	I_g	I_{bn}	I_{wl}	I_{par}	δ			0.0013	46					
	C8	K_t	I_0	I_g	I_{bn}	I_{wl}	I_{par}	δ	I_{incl}			0.0011		101			
	C9	K_t	I_0	I_g	I_{bn}	I_{wl}	I_{par}	δ	I_{incl}	tp		0.0012		56			

Os indicadores das combinações da técnica RNA/MLP são apresentados na Tab. 5 com os melhores resultados experimentados na parametrização e seus indicadores estatísticos após treinamento e validação realizados pelas redes MLP.

Tabela 5 - - Indicadores e configurações do ME e MLP

Modelo	Combinação	Neurônios camada 1	RMSE	rRMSE	MBE	rMBE	R ²	r
Estatístico	C1	-	1.43E-02	30.7015	-7.92E-05	-0.0171	0.79	0.89
MLP	C1	25	1.40E-01	30.1568	-008E-03	-1.7499	0.80	0.89
MLP	C2	25	1.14E-01	25.0759	003E-03	0.6128	0.86	0.93
MLP	C3	25	1.14E-01	24.9003	008E-03	1.7263	0.86	0.93
MLP	C4	25	3.79E-02	8.0767	549E-06	0.1168	0.99	0.99
MLP	C5	25	3.89E-02	8.5211	829E-06	0.1815	0.98	0.99
MLP	C6	25	3.80E-02	8.1979	-479E-06	-0.1033	0.99	0.99
MLP	C7	25	3.66E-02	7.8680	002E-03	0.3456	0.99	0.99
MLP	C8	25	3.27E-02	7.1144	867E-06	0.1884	0.99	0.99
MLP	C9	25	3.52E-02	7.6692	328E-06	0.0713	0.99	0.99

Observa-se nos indicadores de (K_d) que o índice de correlação (r) entre os valores medidos e estimados nas bases de validação variou entre 0,89 a 0,99 mostrando a princípio uma boa correlação.

Os índices de desempenho (rRMSE) da RNA/MLP variou entre os valores 30.167% a 7.11% entre todas as combinações. A rede apresentou os melhores resultados de validação na base quando 8 variáveis de entrada (combinação 8) participaram do treinamento, e quando foram utilizadas apenas uma (1) camadas escondidas.

Analisando os indicadores da Tab. 5, a RNA/MLP apresenta um desempenho melhor que (ME) em todas as combinações, e em média uma redução de aproximadamente 10,10% do coeficiente de correlação mostrando como isso que a inserção de outras variáveis no treinamento contribuiu para um melhor desempenho.

Para um melhor entendimento de exatidão entre os valores medidos e estimados, a Fig. 3 mostra a dispersão apresentada com todas as combinações com a RNA/MLP desde C1 até C9 e sua melhor combinação para estimativa da fração da irradiação difusa em C8.

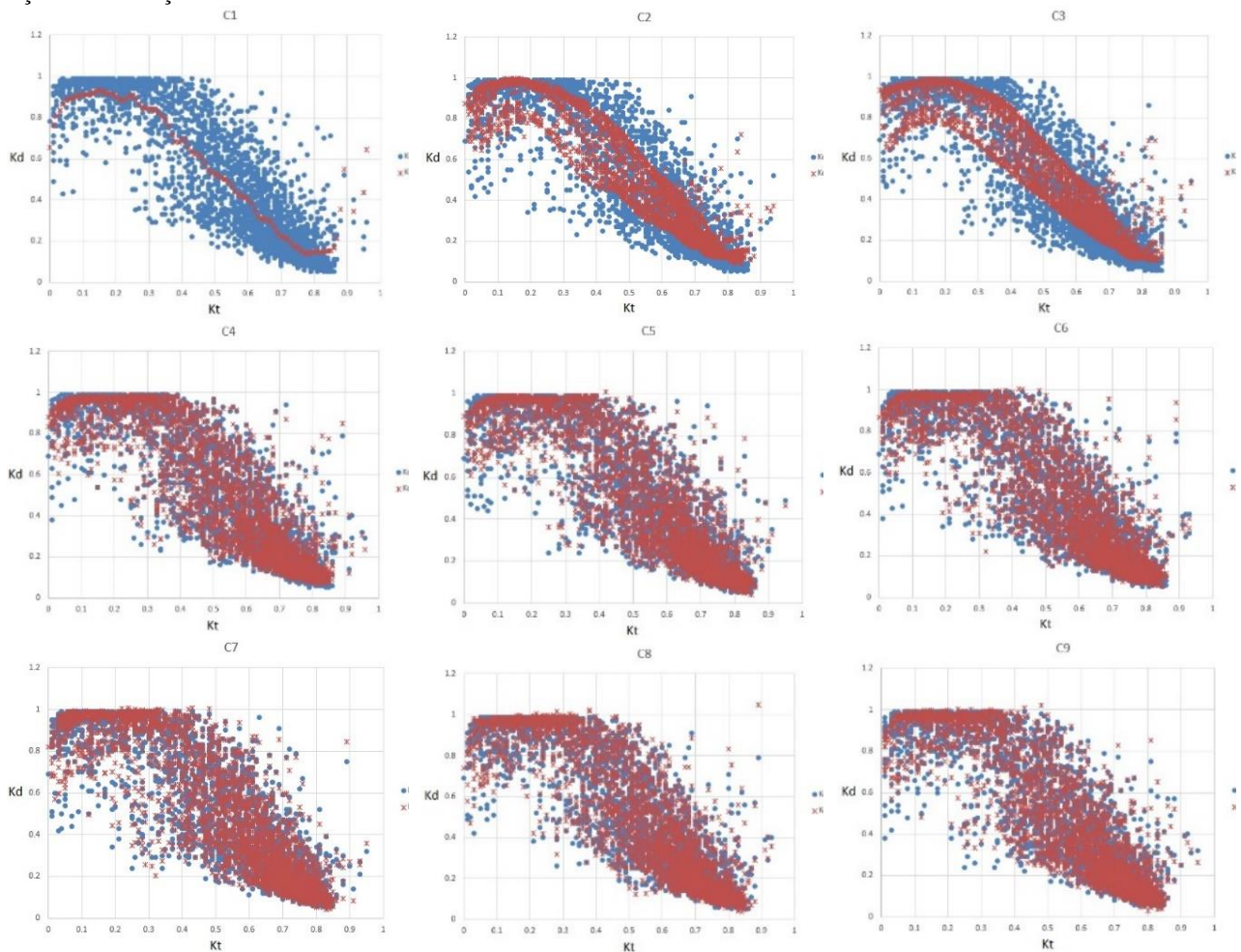


Figura 3: Curva de dispersão de K_d x K_T - de todas as combinações (ponto azul – valor medido/ponto marrom valor estimado).

Importante ressaltar que os resultados encontrados se referem 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 da rede.

Dependendo do algoritmo de treinamento, parâmetros como a taxa para aprendizagem e do termo momentum visam acelerar a convergência ao menor erro entre o valor medido e estimado.

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 serão necessárias outras parametrizações como na escolha do número de camadas, de neurônios e da função de ativação sendo também possível avaliar se a inserção de cada variável irá impactar no desempenho da rede com valores mais significativos para estimativa de K_d .

Para finalizar são apresentados na Tab. 5 alguns índices estatísticos de outras latitudes como indicadores de coeficiente de correlação (r) e/ou do índice de desempenho (RMSE) de alguns trabalhos da literatura que utiliza a RNA/MLP e o melhor resultado da rede para partição horária encontrado neste trabalho, porém não com as mesmas variáveis de entrada.

Tabela 5 – Índices estatísticos de outras latitudes e para Campo Mourão.

Autores	Localidade	Técnicas	Variáveis de entrada	r	RMSE
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	0.9340	
Ihya et. al (2014)	Marrocos	Perceptron	Fração da irradiação global e fotoperíodo	0.9596	18.73
Jiang (2008)	Zhengzhou, China	Perceptron	Índice de claridade	0.9591	
Técnica/MLP	Campo Mourão/Pr	Perceptron	Fração da irradiação global, Radiação no topo da atmosfera, Global na superfície e inclinada, de Ondas longas e PAR, declinação solar e temperatura.	0,99	7.11

Com os resultados observa-se que a rede MLP proporcionou um melhor desempenho na estimativa da fração da irradiação difusa (K_d) com o uso de variáveis climatológicas e meteorológicas, como também melhor que o modelo estatístico e dos apresentados em outras localidades.

5 CONCLUSÕES

Os resultados apresentados nos mostram que o modelo estatístico pode ser utilizado para estimativa da fração transmitida da irradiação solar difusa (K_d) com uma equação de estimativa similar aos apresentados na literatura com um coeficiente de determinação ($R^2 = 0.79$) mesmo com as diferenças das variações astronômicas e climáticas de cada localidade. O valor encontrado do coeficiente de determinação R^2 denota que existe uma correlação entre a fração da irradiação solar difusa e irradiação solar global. Com a utilização RNA/MLP observou-se uma expressiva melhora em comparação ao modelo estatístico. Apesar dos modelos estatísticos apresentados na literatura terem apenas uma variável dependente (K_T) não se pode afirmar que apenas esta variável influencia a estimativa de K_d . As diversas condições atmosféricas, geográficas e meteorológicas que acontecem ao longo do dia como a nebulosidade, a humidade, a temperatura, além de outros constituintes com poeiras e partículas em suspensão na atmosfera, podem também influenciar a variabilidade de K_d . O resultado mostrou que existe a possibilidade de se obter indicadores melhores com mais de uma variável, e que K_d também está fortemente correlacionado com outras variáveis climatológicas. Por fim, outro fato relevante é que os resultados podem influenciar diretamente os estudos e os projetos, como por exemplo nos investimentos de uma nova planta fotovoltaica, no desenvolvimento de um novo cultivar e no controle de constituintes atmosféricos.

Por fim e para estudos futuros pretende-se utilizar a rede treinada em outras localidades com latitudes diferentes utilizando as mesmas variáveis participantes do treinamento e da validação.

Agradecimentos

Os autores agradecem a UTFPR pelo apoio e infraestrutura disponibilizada para o desenvolvimento desta pesquisa e a COPEL-Distribuição pelo apoio e financiamento dos recursos para realização deste projeto de P&D "ANEEL PD 28660464/2017 - Metodologia Para Análise, Monitoramento e Gerenciamento da GD por Fontes Incentivadas".

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.
- 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.
- Chandrasekaran, J.; Kumar, S.. Hourly diffuse fraction correlation at a tropical location. *Solar Energy*, vol. 53, no 6, p. 505 – 510, 1994.
- 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.
- 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.
- Escobedo, J. F.; Dal Paí, A.; Lenz, L.C.; Prado, M. B.. Efeito da cobertura do céu no aumento das frações solar difusa e par-difusa em estufa de polietileno. *Avances en Energías Renovables y Medio Ambiente*, Vol. 17, pp.02.19-02.25, 2013.
- 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.
- 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.
- 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.
- Khatib, T.; Mohamed, A.; Sopian, K.; Mahmoud, M.. Assessment of Artificial Neural Networks for Hourly Solar Radiation Prediction. *International Journal of Photoenergy*, vol. 2012, p. 1 - 8, 2012.
- Khorasanizadeh, H.; Mohammadi, K.; Goudarzi, N.. Prediction of horizontal diffuse solar radiation using clearness index based empirical models: A case study. *International Journal of Hydrogen Energy*, vol. 41, p. 21888 – 21898, 2016.
- 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. 2014
- 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.
- 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.
- 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.; Paliatatos, 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, 2002.

- 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.
- Paulescu, E.; Blaga, R.. Regression models for hourly diffuse solar radiation. *Solar Energy*, vol. 125, p. 111 – 124, 2016.
- Reindl, D. T.; Beckman W. A.; Duffie, J. A.. Diffuse fraction correlations. *Solar Energy*, v.45, p.1 – 7, 1990.
- Ridley, B.; Boland, J.; Lauret, P..Modelling of diffuse solar fraction with multiple predictors. *Renewable Energy*, vol. 35, p. 478–483, 2010.
- 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.
- 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
- Stone, R.J.. Improved statistical procedure for the evaluation of solar radiation estimation models. *Solar Energy*, vol. 51, no 4, p. 289 – 291, 1993.
- Teke, A.; Yıldırım, H. B.; Celik, O.. Evaluation and performance comparison of different models for the estimation of solar radiation. *Renewable and Sustainable Energy Reviews*, vol. 50, p. 1097 – 1107, 2015.
- 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.
- Willmott, C. J. On the validation of models. *Physical Geography*, vol. 2, p. 184-194, 1982.
- Xue, Xinhua. Prediction of daily diffuse solar radiation using artificial neural networks. *International Journal of Hydrogen Energy*. vol. 42, p. 28214 – 28221, 2017.
- 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.

Machine learning model to estimate the diffuse fraction of global radiation based on climatological variables for a subtropical climate zone with atlantic forest biome, mesothermal humid.

Abstract. The work proposes the use of a Machine Learning Technique to estimate the diffuse fraction (K_d) of global solar irradiation from a set of combinations of measured and calculated astronomical and meteorological variables that were extracted from 2020 to 2022 from a of the six Solar Energy Research Stations of the Solar Energy Laboratory – Labens of the Federal Technological University of Paraná. The machine learning technique used is Artificial Neural Networks (RNA/MLP) and to verify the performance of the technique, a statistical model (ME) is created using only the transmitted fraction of global solar irradiation (K_T) as a reference variable. The input variables for ANN/MLP training comprise values of nine measured and calculated astronomical and meteorological variables. The performance is evaluated by the correlation coefficient (r) and the precision index (RMSE) and the statistical model (ME) developed and adjusted by 4th order polynomial regression and presents a coefficient of determination (R^2) of 0.7919, which shows that K_d is correlated with K_T statistically. The performance index of the statistical model (ME) in validation with the database was RMSE = 30.7015%, which shows a good approximation of the estimated values to the measured values and a correlation of $r=0.8899$, which shows the importance from independent variables to dependent variable. It can be seen in the results that with the ANN/MLP the performance indicators (RMSE) showed a reduction in relation to the ME by approximately 23%, which shows that the insertion of variables contributes to the estimates of (K_d) with the astronomical conditions, meteorological and geographic used.

Keywords: Diffuse irradiation, Artificial neural networks, Multilayer Perceptron.