

LISETT ROCIO ZAMORA ORTEGA

**ESTIMATIVA DA RADIAÇÃO SOLAR GLOBAL PELOS MODELOS DE
HARGREAVES E APRENDIZADO DE MÁQUINA EM 11 REGIÕES DE SÃO
PAULO /BRASIL**

Botucatu

2020

LISETT ROCIO ZAMORA ORTEGA

**ESTIMATIVA DA RADIAÇÃO SOLAR GLOBAL PELOS MODELOS DE
HARGREAVES E APRENDIZADO DE MÁQUINA EM 11 REGIÕES DE SÃO
PAULO /BRASIL**

Dissertação apresentada à Faculdade de Ciências Agronômicas da Unesp Câmpus de Botucatu, para obtenção do título de Mestre em Agronomia (Irrigação e Drenagem).

Orientador: Prof. Dr. João Francisco Escobedo

Botucatu

2020

O77e

Ortega, Lisett Rocio Zamora

Estimativa da radiação solar global pelos modelos de Hargreaves e aprendizado de máquina em 11 regiões de São Paulo/Brasil / Lisett Rocio Zamora Ortega. -- Botucatu, 2020
63 p. : il., tabs., mapas

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrônômicas, Botucatu
Orientador: João Francisco Escobedo

1. Radiação solar. 2. Modelo de Hargreaves-Samani. 3. Técnicas de aprendizado de máquinas. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrônômicas, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: ESTIMATIVA DA RADIAÇÃO SOLAR GLOBAL PELOS MODELOS DE HARGREAVES E APRENDIZADO DE MÁQUINA EM 11 REGIÕES DE SÃO PAULO/BRASIL

AUTORA: LISETT ROCIO ZAMORA ORTEGA

ORIENTADOR: JOÃO FRANCISCO ESCOBEDO

Aprovada como parte das exigências para obtenção do Título de Mestra em AGRONOMIA (IRRIGAÇÃO E DRENAGEM), pela Comissão Examinadora:

Prof. Dr. JOÃO FRANCISCO ESCOBEDO
Engenharia Rural / Faculdade de Ciências Agrônomicas de Botucatu

Prof. Dr. ENZO DAL PAI
Pós-Doutorando / Engenharia Rural / Faculdade de Ciências Agrônomicas de Botucatu

Prof. Dr. ADRIANO DE SOUZA MÂRQUES
Departamento de Engenharia da Computação / Instituto Federal de Educação, Ciência e Tecnologia de São Paulo / Birigui

Botucatu, 17 de fevereiro de 2020

Dedico a minha grande e
maravilhosa família. Juana,
Cirilo e Marcelo que me
ajudaram em todo momento. Ao
professor João Francisco, que a
pesar de todo sempre acreditou em
mim.

AGRADECIMENTOS

A Deus.

A meu orientador, **Prof. Dr. João Francisco Escobedo**, pela paciência, compreensão e oportunidade de trabalhar e pelo apoio durante o mestrado.

Aos meus queridos pais Cirilo e Juana pelo apoio incondicional.

A meu irmão Marcelo sua esposa Karen e meus sobrinhos que independente da distância sempre torceram por mim.

A minha melhor amiga Wendy Obrigado por entender-me, chatices e por estar sempre me apoiando.

Aos funcionários do Departamento de Engenharia Rural da Faculdade de Ciências Agronômicas- FCA.

A Fran por sempre me motivar com o trabalho, a Profa. Alba por sempre se preocupar por mim a Tami que foi de muito apoio sempre.

Aos amigos do grupo Maurício, Taiza, Edson, Jancer, Angélica. Por amizade e experiências compartilhadas.

O Presente trabalho foi realizado com apoio da coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES)- código de financiamento 001

RESUMO

No presente trabalho é descrito o estudo comparativo de métodos de estimativas da irradiação solar global (H_G) diária através do modelo de Hargreaves-Samani (H-S) $H_G/H_0 = a \Delta T^{0.5}$ e duas técnicas de Aprendizado de Máquina (AM), Máquinas de Vetores de Suporte (MVS) e Redes Neurais Artificiais (RNA). A base de dados utilizada foi obtida em 11 cidades do estado de São Paulo de diferentes classificações climáticas no período de 2013-2017. Por meio de regressão entre a transmissividade atmosférica (H_G/H_0) e raiz quadrada da diferença de temperatura ($\Delta T^{0.5}$). O modelo estatístico H-S foi calibrado e determinado para os valores da constante (a) e equações que permitem estimar H_G com baixos coeficientes de determinação para duas condições: 11 cidades individualmente e total. Os modelos de H-S foram validados por meio de correlações entre os valores estimados e medidos através dos indicadores de correlação (r) e $rRMSE$ cujos valores indicaram que os modelos podem estimar H_G com razoável precisão e exatidão. As técnicas computacionais, MVS e RNA, foram treinadas com 70% dos dados nas mesmas variáveis usadas no modelo de H-S, e posteriormente foram treinadas com entradas de mais 4 variáveis meteorológicas totalizando 5 combinações. Os treinos foram validados usando uma base de dados independente de 30% da base. Os indicadores estatísticos (r) das correlações mostraram que o modelo H-S pode estimar H_G com baixos coeficientes de determinação. Os indicadores estatísticos $rMBE$, MBE , $rRMSE$, $RMSE$ indicam que o modelo H-S pode ser utilizado na estimativa de H_G com razoável exatidão e precisão. Os indicadores estatísticos obtidos pelos 5 combinações das técnicas MVS e RNA (diária) indicaram que os modelos podem ser utilizadas nas estimativas de H_G com elevadas correlações, precisão e exatidão. A comparação dos indicadores estatísticos (r), $rMBE$, MBE , $rRMSE$, $RMSE$, entre o modelo de H-S e os modelos de redes MVS e RNA combinação 5, mostrou que: a técnica MVS possui melhor desempenho que a técnica RNA e o modelo estatístico H-S; e o modelo de RNA obteve melhores resultados que o modelo de H-S, nas estimativas de H_G .

Palavras-chave: Radiação solar, Modelo de Hargreaves-Samani, Técnicas de Aprendizado de Maquinas.

ABSTRACT

This work describes the comparative study of methods for estimating daily global solar irradiation (H_G) using the Hargreaves-Samani (H-S) model $H_G / H_0 = a \Delta T^{0.5}$ and two Machine Learning techniques (AM), Support Vectors Machines (MVS) and Artificial Neural Networks (ANN). The database used was obtained in 11 cities by the state of São Paulo from different climatic classifications between period 2013-2017. Through regression between different atmospheric transmissivity (H_G / H_0) and square root of the temperature difference ($\Delta T^{0.5}$). The H-S statistical model was calibrated and determined for the values of constant (a) and equations that allow estimating H_G with low determination coefficients for two conditions: 11 cities individually and total. The H-S models were validated by correlations between the estimated and measured values using the correlation indicators (r) and rRMSE, whose values indicated that the models can estimate H_G with reasonable precision and exactitude. The computational techniques, MVS and RNA, were trained with 70% of the data in the same variables used in the H-S model, later they were trained with inputs of 4 more meteorological totalling 5 combinations. The training was validated using an independent database of 30%. The statistical indications of the correlations showed that the H-S model can estimate H_G with low determination coefficients. The statistical indications rMBE, MBE, rRMSE, RMSE indicate that the H-S model can be used to estimate H_G with reasonable accuracy and precision. The statistical indications obtained by the 5 combinations of the MVS and RNA techniques (daily) indicated that the models can be used in the estimates of H_G with high correlations, precision and accuracy. The comparison of the statistical indicators (r), rMBE, MBE, rRMSE, RMSE, between the HS model and the MVS and RNA combination 5 network models, showed that: the MVS has better performance than the RNA technique and the statistical model HS; and the RNA model obtained better results than the H-S model, in the H_G estimates.

Keywords: Solar Radiation, Hargreaves-Samani Model, Machine Learning Techniques.

LISTA DE TABELAS

	Página
Tabela1- Informações das regiões geográficas.....	31
Tabela 2- Dados para elaboração e validação do Modelo Estatístico.....	31
Tabela 3- Dados para elaboração e validação das Redes.....	34
Tabela 4- Arquitetura das combinações.....	35
Tabela 5- Escala classificatória dos índices estatísticos.....	37
Tabela 6- Regiões Geográficas e coeficiente de correlação a de H-S.....	40
Tabela 7- Localidades Altitudes Clima e Coeficiente de H_G	41
Tabela 8- Coeficiente de H-S, Desvio Padrão e Porcentagem.....	42
Tabela 9- Valores dos indicadores estatísticos MBE,e RMSE.....	44
Tabela 10- Coeficientes de parametrização das combinações 1 e 5.....	47
Tabela 11- Combinação 1 para os modelos da redes de MLP e MVS.....	53
Tabela 12- Combinação 5 para os modelos da redes de MLP e MVS	54
Tabela 13 Comparação do desempenho dos modelos com resultados da literatura.....	56
Tabela 14- Comparação do desempenho dos modelos.....	58

LISTA DE FIGURAS

	Página
Figura 1- Modelo do Neurônio Artificial.....	28
Figura 2 - Arquitetura da Rede Neural Perception Multicamadas.....	29
Figura 3 - Localização das 11 regiões geográficas.....	30
Figura 4 - Diagrama em blocos demonstrando a arquitetura da RNA.....	33
Figura 5 - Correlações da irradiação global diária.....	38
Figura 6 - Curvas de regressão linear para o modelo HS das 11 cidades.....	39
Figura 7- Gráficos de coeficiente de H-S das 11 cidades.....	42
Figura 8- Comparação entre os valores estimados de H_G pelas equações.....	43
Figura 9- Correlações da irradiação global diaria total.....	45
Figura 10- Comparação entre os valores estimados de H_G	46
Figura 11- Diagramas de dispersão entre os valores estimados de H_G , e pelos modelos gerados pelas RNA e MVS.....	48
Figura 12- Indicativos estatísticos rMBE, rRMSE na validação: combinação 1 e 5.....	51
Figura 13- Desempenho do modelo de redes Total da combinação 1 e 5.....	55

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquinas
ME	Modelo Estatístico
H-S	Hargreaves-Samani
RNA	Redes Neurais Artificiais
MLP	Rede Neural Preceptor Multicamadas
MVS	Máquinas de Vetores de Suporte
H_0	Irradiação no topo da atmosfera (MJ/m^2)
H_G	Irradiação global na superfície horizontal (MJ/m^2)
T_{MAX}	Temperatura máxima
T_{MIN}	Temperatura mínima
T_{MED}	Temperatura media
P	Precipitação
U2	Velocidade do vento
UR	Umidade relativa
RMSE	Raiz Quadrada do Erro Quadrático Médio (%)
MBE	Erro Quadrático Médio
R^2	Coeficiente de determinação
r	Coeficiente de correlação
ΔT	Diferencial de temperatura
INMET	Instituto nacional de meteorologia

SUMARIO

1 INTRODUÇÃO	23
2 REVISÃO DE LITERATURA	26
2.1 Modelo de Hargreaves e Samani	26
2.2 Aprendizado de maquinas(AM)	26
2.2.1 Aprendizado de maquinas de vetores de suporte(MVS)	27
2.2.2 Redes neural artificial (.RNA)	27
3 MATERIAL E MÉTODOS	30
3.1 Localização e clima	30
3.2 Modelo Estadístico	31
3.3 Treinamento de MVS E MLP	32
3.3.1 Aplicação da Máquina Vetor de Suporte (MVS)	32
3.3.2 Aplicação da Rede Multilayer Perceptron (MLP)	33
3.4 Geração de modelos da estimativas com as técnicas de aprendizado de maquinas MVS e RNA	34 36
3.4.1 Indicativo Estatísticos	
4. RESULTADOS E DISCUSSÕES	38
4.1. Modelo Estatístico	38
4.1.1 Obtenção e validação das equações de H-S para as 11 cidades, localização e clima	38
4.1.2 Obtenção e validação da equação H-S geral ou espacial	45
4.2 Treinamento e validação das técnicas RNA E MVS	46
4.2.1 Treinamento e Validação das Técnicas MLP e MVS	46
4.2.2 Desempenho do modelo de redes total (r e rRMSE)	54
4.2.3 Comparação do desempenho do modelo MVS e RNA de com resultados da literatura	55
4.3. Comparação do desempenho entre os modelos estatístico e as redes MVS e RNA	57
5. CONCLUSÕES	59
REFERÊNCIAS	61

1 INTRODUÇÃO

A radiação solar global H_G é uma das principais fontes de energia para os processos biológicos, físicos e químicos na superfície terrestre, e o conhecimento de H_G , por meios de medidas ou modelos, são de fundamental importância para elaboração de projetos otimizados nas áreas climáticas; energias renováveis obtidas por processos térmicos e fotovoltaicos com coletores solares; na arquitetura e agricultura em modelos computacionais de crescimento e produtividade de cultura agrícolas e estimativas de evapotranspiração. O Brasil, país com dimensão continental, possui um déficit de informações e disponibilidade de informações não só de H_G , como também de outras radiações importantes nas aplicações citadas como as radiações difusa, direta, PAR e as radiações espectrais entre outras. (Silva, 2008). Com necessidade de suprir a falta dessas informações, atualmente, diversos pesquisadores vêm desenvolvendo modelos estatísticos (modelos empíricos), físicos e por técnicas de Aprendizado de Máquina (AM), que permitem estimar valores de H_G por meio de variáveis mais facilmente monitorada nas estações meteorológicas. Os modelos estatísticos, devido a sua simplicidade e precisão, normalmente, são recomendados para os locais em que os modelos são gerados ou para regiões de clima semelhante. Os modelos físicos (transferência radiativa) são mais complexos que os modelos estatísticos e exigem medidas com precisão de variáveis meteorológicas na superfície e na atmosfera, que por essa razão existem poucos estudos disponíveis. Os modelos de AM utilizando diferentes variáveis de entrada são capazes de solucionar problemas complexos, gerando estimativas das radiações com maior precisão.

O principal modelo é o de Angstrom-Prescott que estima H_G a partir do brilho solar, a qual é monitorada mais comumente nas estações meteorológicas. Para locais, onde brilho solar não é monitorado, outro modelo é usado: Hargreaves e Samani (1982) o qual que usa a temperatura do ar como variável de entrada. Esse modelo apesar de ser muito limitado em função da baixa precisão, ainda é muito utilizado em locais que só medem a temperatura, como em regiões carentes ou países pobres. Há também outros modelos, denominados de modificados que associam a temperatura do ar à outras variáveis de entrada como precipitação, altitude, entre outras como mostram os trabalhos de Annandele.

(Almorox, 2011; Bechini, 2000; Bojanowski, 2013; Donatelli, 2013; Chen e Zhao, 2004; Hansen, 1999; Hunt, 1998; Podestá e Skansi, 2004; Weiss e Hays, 2004). As técnicas de AM possuem capacidade computacional adquirida por meio de aprendizado e generalização. Tal habilidade permite, sem a necessidade de conhecer o objeto investigado, modelar um sistema conhecendo apenas os valores de entrada e saída. A maior vantagem do uso de modelos baseados em AM é a sua simplicidade, otimização do tempo e esforço computacional requerido após a fase de treinamento (Soares & Da Silva, 2011). Entre as diferentes técnicas de AM destacam-se: Máquina Vetor de Suporte (MVS), Algoritmos Genético (AG), Programa de Expressão do Gene (PEG), Redes Neurais Artificiais (RNA), Sistema de Inferência Neuro-Fuzzy (SINF).

A técnica de Aprendizado de Máquina (AM) tem ganhado destaque na geração de estimativa de H_G e de outras radiações com resultados satisfatórios estatisticamente desde os estudos iniciais realizados por Elizondo et al. (1994) e posteriormente por meio de: (Bocco, 2006; Da Silva, 2017; Dos Santos, 2016; Kumar, 2015; Landeras, 2012; Lyra et al., 2016; Ramedani, 2014). Existem poucos trabalhos disponíveis na literatura sobre modelos AM comparado aos modelos estatístico de Hargreaves e Samani para estimativa de H_G . Resultados obtidos mostram que o desempenho dos modelos de aprendizados de maquina AM é superior aos estatísticos.

Diante da necessidade de mais estudos comparativos entre esses dois modelos, no presente trabalho é realizada uma análise do desempenho dos modelos estatístico Hargreaves e Samani e de técnicas de Aprendizado de Máquina (AM): Máquina Vetor de Suporte (MVS) e Redes Neurais Artificiais (RNA) para 11 locais ou cidades com diferentes classificações climáticas do estado de São Paulo com os seguintes objetivos:

- 1- Calibrar o modelo de H-S para 11 cidades com diferentes classificações climáticas determinando-se o valor da constante (a) da equação clássica de H-S o coeficiente de determinação das equações H-S.

2- Calibrar o modelo de H-S para uma base de dados total, determinando se o valor da constante (a) da equação clássica de H-S geral e o coeficiente de determinação das equações H-S geral.

3- validar as equações de H-S para 11 cidades com diferentes classificações climática e total.

4- Efetuar o treinamento das redes MVS e MLP para cinco combinações de entradas de variáveis meteorológicas.

5- Efetuar o treinamento das técnicas MVS e MLP para as diferentes combinações de entradas de variáveis meteorológicas para 11 cidades e total através dos indicadores estatísticos coeficiente de correlação e RMSE.

6- Efetuar a validação das redes MVS e MLP para as diferentes combinações de entradas de variáveis meteorológicas para a base de dados e total, através dos indicadores estatísticos coeficiente de correlação e RMSE.

7- Comparar o desempenho do modelo estatístico as equações de H-S para 11 cidades climática e total, e do modelo de redes MVS e MLP para as combinações de entradas de variáveis meteorológicas (a primeira e a de melhor desempenho) para 11 cidades e total.

2 REVISÃO DE LITERATURA

2.1 Modelo de Hargreaves e Samani

A primeira equação de Hargreaves exige apenas dados medidos de temperatura do ar e radiação no topo da atmosfera a mesma tem sido amplamente utilizada por sua simplicidade e potencial recomendado no boletim da FAO-56 para ser usado quando os dados de radiação solar estão ausentes ou de qualidade duvidosa, a equação (1) introduzida por Hargreaves e Samani.

$$H_G = a * \sqrt{\Delta T} * H_0 \dots \dots \dots (1)$$

Onde:

- H_G (MJ / m² dia) é a irradiação solar estimada
- H_0 (MJ / m² dia) é a radiação solar no topo da atmosfera ,
- $\Delta T = (T_{\max} - T_{\min})$ $T_{\max}$ (°C) é a temperatura máxima diária, $T_{\min}$ (°C) é a temperatura mínima diária.

a (°C)^{-0.5} é um coeficiente empírico. Inicialmente, a foi definido para 0,17 para regiões áridas e semiáridas. Hargreaves Samadi recomenda para usar $a = 0,16$ para regiões interiores e $a = 0,19$ para Regiões costeiras. A equação é baseada na suposição de que o diferença entre as temperaturas máximas e mínimas diárias fornece uma indicação geral de nebulosidade.

2.2 Aprendizados de Maquinas (AM)

O princípio de funcionamento das Técnicas de Aprendizado de Máquinas é similar ao processo de aprendizado do cérebro humano. A partir de um conjunto de dados ou de experiências adquiridas ao longo do tempo é possível obter uma série de resultados. Isso se deve ao fato de que as técnicas podem reconhecer padrões de diferentes grandezas, como dados meteorológicos ou climatológicos e através de um algoritmo estabelecer uma associação entre dados de entradas e dados de saídas para adquirir “conhecimento” de forma que produz resultados coerentes em reconhecimentos de padrões ou classificação de dados. (Lorena et al.,2011).

Entre as Técnicas de Aprendizado de Máquinas (AM) utilizados para estimar radiação solar em superfície horizontal ou inclinadas destacam-se as Redes Neurais Artificiais (RNA) amplamente utilizadas na literatura, Máquinas de

Vetores de Suporte (MVS) que vem se destacando com uma alternativa, além de outras como os de Inferência Adaptativo Neuro –Fuzzy (ANFIS) (Yadav e chandel, 2014;Lauret et al., 2015; Santos et .al, 2016; silva et. al, 2017; Assouiline et. al, 2017).

2.2.1 Máquinas de Vetores de Suporte (MVS)

A técnica MVS foi desenvolvida para resolver problemas de classificação, regressão e estimativas com grande capacidade de generalização e introduzida por Vapnik (1995). A MVS é baseada no princípio da minimização do risco estrutural, que tenta minimizar um limite superior de erro de generalização em vez de minimizar o erro de treinamento local. (Cortes e Vapnik, 1995).O conceito básico da técnica é realizar a estimativa da rede através de algoritmos de otimização matemática envolvendo poucos parâmetros de ajuste em seus algoritmos de treinamento que depende da sua utilização e consequentemente do algoritmo de treinamento a ser utilizados (Silva et al., 2017).

Na maquina RBF, diferentemente do kernel linear, e possível resolver problemas, originalmente, não linearmente separáveis, através do mapeamento para um espaço de maior dimensão. Existem dois parâmetros que podem ser variados em busca de um melhor resultado para o aprendizado do classificador, são eles: γ (gamma) e C (custo).

No kernel RBF o numero de funções radiais e os seus respectivos centros são definidos pelos vetores suporte obtidos; outro motivo para uso do kernel RBF e o baixo numero de parâmetros (dois) que influenciam o resultado do aprendizado.

2.2.2. Rede Neural Artificial (RNA)

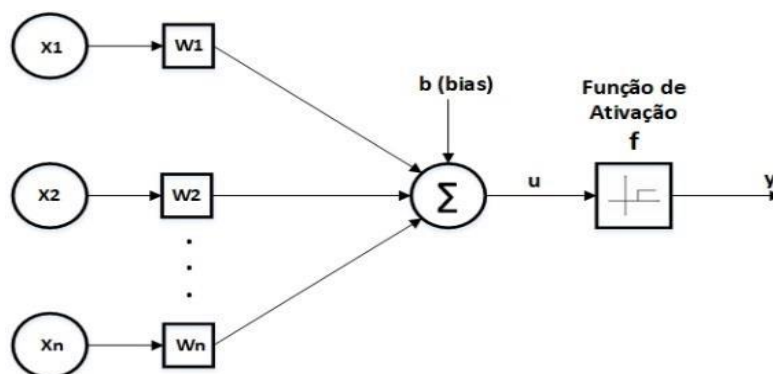
A Rede Neural artificial (RNA) apresentam uma estrutura formada por um conjunto de elementos chamados neurônios distribuídos paralelamente. É composta por uma camada de entrada e uma de saída conectada por uma ou mais camadas ocultas. O processamento de informações entre as camadas é realizado por um algoritmo capaz de analisar funções complexas e não lineares os que nem sempre é possível com o uso de técnicas estatísticas convencionais.

São utilizadas para o reconhecimento e classificação de padrões, agrupamento, previsão e estimativas. (Lorena et al., 2011).

Entre os diferentes tipos de RNA existente, destaca-se a do tipo *Multilayer Perceptron* (MLP), a RNA é usada com sucesso na resolução de problemas complexos que incluem o reconhecimento de padrões, identificação, classificação e regressão. Sua estrutura não linear consegue captar características mais complexas dos dados, o que nem sempre é possível com a utilização das técnicas estatísticas tradicionais (Bocco et al., 2006; Soares e Da Silva, 2011).

Na Figura 1 apresenta a estrutura básica de um neurônio artificial formado por um conjunto de conexões (*sinapses*) que recebem os valores de entrada ($x_1, x_2, \dots, x_i$) definidos pelos pesos sinápticos ($w_1, w_2, \dots, w_i$). Através de um combinador linear (Σ), executa a soma ponderada das entradas, produzindo uma saída chamada de potencial de ativação (u) que é então 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) (Marzouq et al. 2017).

Figura 1- Modelo do Neurônio Artificial.



Fonte: McCullochpitts(1943).

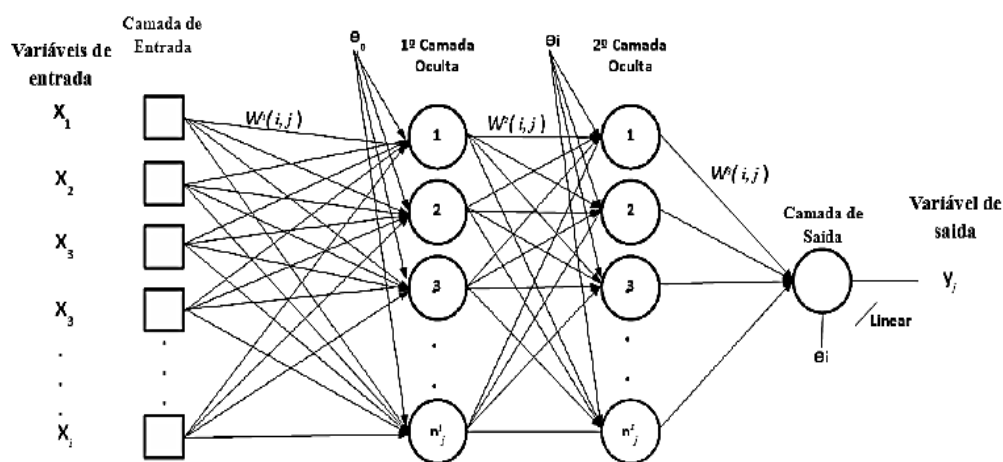
A principal característica de uma RNA está na capacidade de aprendizado proporcionado pela arquitetura e no processo de treinamento utilizado. O aprendizado é consequência das mudanças de valores dos pesos sinápticos, conectados aos neurônios entre as camadas com o objetivo de aproximar os valores da saída da rede (y) com os valores desejados. Esse processo é realizado por um algoritmo de treinamento que define aspectos importantes como o número

de neurônios, número de camadas ocultas, função de ativação e conjunto de dados para treinamento e validação.

A rede Multilayer Perceptron (MLP). São utilizadas para resolução de problemas complexos e consiste de um conjunto de neurônios com três ou mais camadas interligadas. São classificadas como redes do tipo *feedforward* (alimentação a frente) de camadas múltiplas. O treinamento é realizado de forma supervisionada, ou seja, para cada informação de entrada tem-se uma respectiva saída desejada (Alam et al. 2009).

A Figura 2 mostra a estrutura da rede MLP, onde o fluxo de treinamento tem início quando as variáveis de entrada (x) são aplicadas na camada de entrada, deslocando-se às camadas ocultas até a camada de saída finalizando o processamento por uma função de ativação (normalmente linear) na camada de saída (Rehman e Mohandes, 2008).

Figura 2 - Arquitetura da Rede Neural Perception Multicamadas.



Fonte: Yadav e Chandel (2014).

Durante o processo de treinamento, algoritmos procuram ajustar os pesos ($w_{i,j}$) entre os neurônios das entradas e a saída desejada. Algoritmos como o *backpropagation* (BP) otimizados por *Levenberg-Marquardt* ajustam os pesos através de taxas de aprendizagem (controle dos bias - Θ) e de taxas de convergências (controle de velocidade) tornando o processo de treinamento da rede mais rápido e eficiente (Linares-Rodriguez et al., 2011; Braga et al. 2012).

3 MATERIAL E MÉTODOS

3.1 Localização e Clima

As 11 regiões geográficas intermediárias do Estado de São Paulo, possui grande gradiente de altitude, entre 337 a 1642 m, esta diferença provoca variações na temperatura do ar. A figura 3 mostra as onze regiões do Estado de São Paulo.

Figura 3- Localização das 11 regiões geográficas.



Fonte: Google Earth

A base de dados climáticos, medida durante o período 2013-2017, foi obtida de 11 Estações Meteorológicas Automáticas (EMA's) em regiões de climas, latitudes e altitudes diferentes. Foram selecionadas 11 cidades interligadas por meio de estações automáticas supervisionadas pelo Instituto Nacional de Meteorologia (INMET). Na tabela 1 são mostradas as informações geográficas das localidades, latitudes, clima, altitudes, e fonte de aquisição dos dados.

Tabela 1 Informações das regiões geográficas, Localidades, Latitudes, longitudes, climas e fontes.

Regiões Geográficas	Cidade	Latitude	Clima	Altitude	Fonte
1.São Paulo	São Paulo	23°30'S	Subtropical úmido	792,06	INMET
2.Sorocaba	Sorocaba	23°21'S	Subtropical	609	INMET
3.Bauru	Lins	21°39'S	Tropical de altitude	459	INMET
4.Marília	Ourinhos	22°57'S	Subtropical	448	INMET
5.Presidente Prudente	Presidente prudente	22°07'S	Tropical	435,55	INMET
6.Araçatuba	Ilha Solteira	20°25'S	Tropical semi-úmido	337	INMET
7.São Jose do Rio Preto	Jales	20°09'S	Tropical	457	INMET
8.Ribeirão Preto	Barreto	20°33'S	Tropical com estação seca	533	INMET
9.Araraquara	São Carlos	21°57'S	Tropical de altitude	863	INMET
10.Campinas	Itapira	22°24'S	Tropical de altitude	633	INMET
11.São Jose dos Campos	Campos do Jordão	22°45'S	Oceânico	1642	INMET

3.2 Modelo Estadístico

Para as estimativas com modelo estatístico foi utilizado a metodologia proposta por Hargreaves e Samani que relaciona (H_o) radiação solar no topo da atmosfera e (ΔT) diferencial de temperatura resultando em equações de estimativa ajustadas por regressão. O modelo estatístico foi elaborado a partir da base de dados reorganizada, retirando-se uma parte dos dados para validação do modelo conforme apresentado na tabela 2.

Para treinamento e validação do modelo, os dados foram reorganizados em duas bases 70% e 30% para (H_G). Esta reorganização foi realizada através de análises estatística.

Tabela 2 - Dados para elaboração e validação do Modelo Estatístico.

(Total de dados) (2013-2017)	Total de dados para elaboração do Modelo Estatístico para cada cidade	
	1146 -70%	
	Total de dados para validação das redes	
	493- 30%	

3.3 Treinamento de MVS E MLP

3.3.1 Aplicação da Máquina Vetor de Suporte (MVS)

A implementação da técnica MVS é interessante pela precisão na estimativa o conjunto de dados foram submetidos para execução no software da *Waiko Environment for Knowledge Analysis - WEKA* (Hall et al., 2009) ferramenta *KDD – Knowledge Discovery in Databases* que agrega uma série de algoritmos de preparação de dados, aprendizagem e de validação de resultados de forma interativa. No treinamento e validação foi utilizando o algoritmo de regressão MVS/ (SMOreg) Otimização Mínima Sequencial é utilizado para regressão. SMOreg é um algoritmo simples que rapidamente soluciona e resolve o problema de menor otimização possível com dois multiplicadores de Lagrange (Platt, 1999; Shevade e Murthy, 2000; Smola e Schölkopf, 2004).

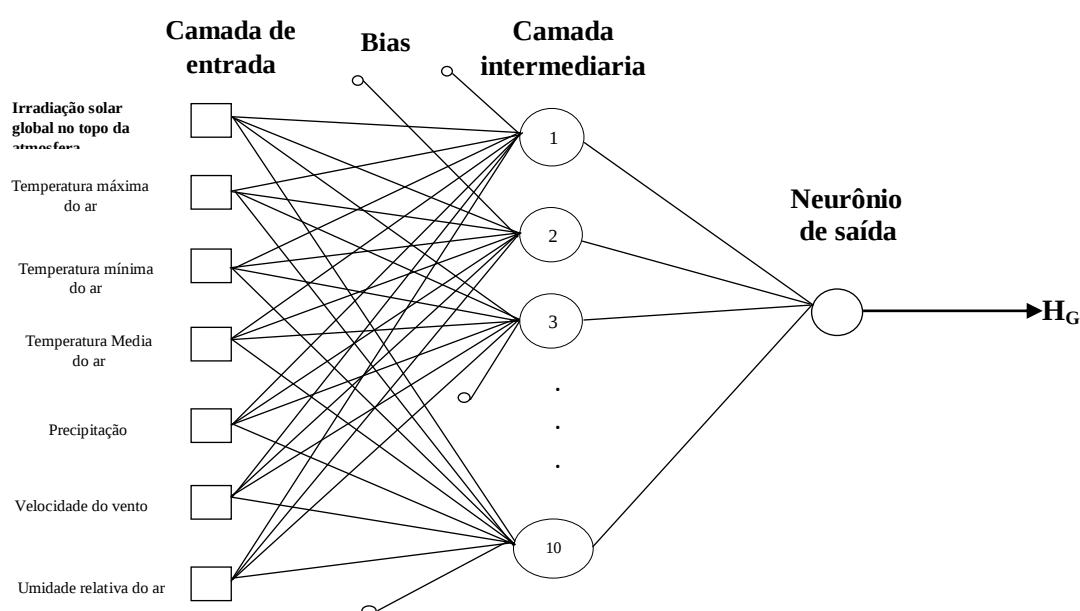
A flexibilidade do SMO na regressão é atribuída ao uso do conjunto de funções de *Kernel*. Pela eficiência computacional, simplicidade e adaptação para otimização de problemas mais complexos, neste estudo com função de transformação de base radial (RBF) de *Kernel* é utilizada para regressão, na aplicação da RBF é necessário ajustar adequadamente os parâmetros C (custo) que determina o ponto de equilíbrio entre a maximização da margem e minimização do erro e que controla a influencia de cada vetor, o parâmetro γ (gamma) que controla a flexibilidade da função de kernel e verifica seu ajustamento overfitting é quando o erro da validação aumentar enquanto o erro do conjunto de treinamento continuar a cair, condição que indica o momento de parar o treinamento e underfitting e o parâmetro ϵ (épsilon) relacionado ao erro de tolerância. Os valores de C e γ são testados e aquele com a melhor precisão na validação são escolhidos.

Os modelos gerados com o algoritmo SMO foram com base na função de transformação de base radial (RBF). Foram testados empiricamente diferentes combinações para os parâmetros obtidos da RBF, depois de inúmeros teste os parâmetros utilizados da RBF foram ($C = 100$, $\gamma = 0,3$ e $\epsilon = 0,001$), que são valores encontrados na literatura. (Mohammadi e Petković, 2015; Mohammadi et al, 2015; Ramedani et al., 2014).

3.3.2 Aplicação da Rede Multilayer Perceptron (MLP)

A Rede *Multilayer perceptron* (MLP) é utilizada para resolução de problemas complexos e consiste de um conjunto de unidades sensoriais com três ou mais camadas interligadas. Na Figura 4 é demonstrada a estrutura de uma MLP, onde a primeira camada é a entrada (variáveis meteorológicas), a segunda possui uma ou mais camadas ocultas de nós computacionais com pesos de ligações e a terceira camada corresponde à saída de nós computacionais. (Bosch et al., 2008; Zanetti et al., 2008; Lyra et al., 2015).

Figura 4- Diagrama em blocos demonstrando a arquitetura da RNA utilizada (MLP).



Fonte: Adaptado de Zanetti et al. (2008).

Para estimativas com a rede MLP foram elaborados uns scripts de treinamento e validação para os conjuntos de dados que foram submetidos no software MATLAB ®2014 que pode ser descrito como uma ferramenta computacional com aplicação em diversas áreas. Permite 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 nas técnicas de aprendizado de máquinas. O treinamento da rede 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. A definição do número de camada escondida (n_1) foi 5, o número máximo de épocas foi fixado em mil (1.000) interações e a função de ativação na camada oculta a tangente Hiperbólica (tansig) definida conforme as características dos dados de entrada e de saída.

São realizados dez (10) treinamentos para cada combinação (tabela 4) mostra as combinações O critério do treinamento é a validação cruzada 70% treinamento, 15% para validação e 15% para teste; assim como no ME para AM a base de dados foi reorganizada a fim de retirar os dados para treinamento e validação do modelo conforme apresentado na tabela 3.

Em cada interações de MBE (Erro quadrático médio) entre os valores estimados e medidos da validação são avaliados e se forem diminuindo significa que a rede está aprendendo Sendo definido a melhor rede em cada treinamento todas as combinações são validadas.

Tabela 3- Dados para elaboração e validação das redes.

(Total de dados) (2013-2017)	Total de dados para treinamento das redes cada cidade
	1146 -70%
(1639-100%) Cada cidade	Total de dados para validação das redes cada cidade
	493- 30%

3.4 Geração de Modelos de estimativas com as técnicas de aprendizado de maquinas MVS E MLP

A correlação entre a transimitividade atmosférica da irradiação solar global (H_G/H_0) com variação da temperatura ($\Delta T^{0.5}$) diária, para a base de dados medida no período de 2013-2017 medidas em as onze cidades do estado de São Paulo, foram treinadas e validadas para as redes MVS e RNA, respectivamente, conforme as 5 combinações citada na tabela de abaixo.

Tabela 4- Arquitetura das combinações geradas pelas técnicas Máquina Vetor de Suporte (MVS) e Redes Neurais Artificiais (RNA) com diferentes variáveis meteorológicas de entrada.

Combinações		
MVS –RFB	RNA MLP	Variáveis Meteorológicas de Entrada
MVS1	RNA1	Ho,Tmax,Tmin
MVS2	RNA2	Ho,Tmax,Tmin,Tmed
MVS3	RNA3	Ho,Tmax,Tmin,Tmed,P
MVS4	RNA4	Ho,Tmax,Tmin,Tmed,P,U2
MVS5	RNA5	Ho,Tmax,Tmin,Tmed,P,U2,Urmed

Na Tabela 4 é apresentada a arquitetura das combinações gerada com as variáveis meteorológicas utilizadas como entrada em cada um das combinações desenvolvida pelas técnicas de AM:

- Combinação 1- irradiação solar global no topo da atmosfera – H_o ($MJ\ m^2\ dia^{-1}$); temperatura máxima - $T(^{\circ}C)$ [máx] e temperatura mínima - $T(^{\circ}C)$ [min].
- Combinação 2- irradiação solar global no topo da atmosfera – H_o ($MJ\ m^2\ dia^{-1}$); e temperatura máxima- $T(^{\circ}C)$ [máx]; temperatura mínima- $T(^{\circ}C)$ [min] e temperatura media $T(^{\circ}C)$ [med].
- Combinação 3- irradiação solar global no topo da atmosfera H_o ($MJ\ m^2\ dia^{-1}$); temperatura máxima $T(^{\circ}C)$ [máx]; temperatura mínima $T(^{\circ}C)$ [min];temperatura media $T(^{\circ}C)$ [med]; e precipitação- P (mm).
- Combinação 4 - irradiação solar global no topo da atmosfera H_o ($MJ\ m^2\ dia^{-1}$); temperatura máxima $T(^{\circ}C)$ [máx]; temperatura mínima $T(^{\circ}C)$ [min];temperatura media $T(^{\circ}C)$ [med]; precipitação- P (mm); e velocidades do vento $U2$ ($m\ s^{-1}$).
- Combinação 5 - irradiação solar global no topo da atmosfera H_o ($MJ\ m^2\ dia^{-1}$);e temperatura máxima $T(^{\circ}C)$ [máx]; temperatura mínima- $T(^{\circ}C)$ [min];temperatura media $T(^{\circ}C)$ [med]; precipitação- P (mm); velocidades do vento $U2$ ($m\ s^{-1}$); e umidade relativa do ar - UR (%)

3.4.1. Indicativos estatísticos

Os índices estatísticos utilizados para avaliar o desempenho de modelos foram: *Mean Bias Error* (MBE), *Relative Mean Bias Error* (rMBE), que fornece informações sobre o desempenho dos modelos a longo prazo, permitindo uma comparação do desvio real entre as estimativas e as medidas, sendo o valor ideal de MBE é “zero”.

Root Mean Square Error (RMSE) e *Relative Root Mean Square Error* (rRMSE), que fornecem informações sobre o desempenho de curto prazo. O valor do RMSE é sempre positivo, quanto menor os valores obtidos para RMSE, melhor desempenho dos modelos (Stone, 1993).

O coeficiente de determinação (R^2) que é uma medida de ajustamento de um modelo estatístico linear generalizado, como a regressão linear, em relação aos valores observados, ele varia entre 0 e 1. O coeficiente de correlação (r) que indica o grau de relacionamento entre as variáveis. Tanto (R^2) e (r), indicam a percentagem, ou seja, o quanto o modelo consegue explicar os valores observados (Stone, 1993). Para análise do *rRMSE* será analisada pela escala classificatória conforme tabela 5. (Li et al., 2013).

Os indicadores são obtidos pelas seguintes relações: onde E_i são os valores estimados, $\bar{E}_i$ é a média dos valores estimados M_i , são os valores medidos, $\bar{M}_i$ é a média dos valores medidos, N é o número de observações.

$$MBE = \left[\frac{\sum_i^N (E_i - M_i)^2}{N} \right]$$

$$RMSE = 100 * \frac{[\frac{\sum_i^N (E_i - M_i)^2}{N}]^{1/2}}{\bar{M}}$$

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

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

Tabela 5- Escala classificatória dos índices estatísticos.

Fraco	Aceitável	Bom	Excelente
RMSE ≥ 30	20% ≤ RMSE < 30%	10% ≤ RMSE < 20%	RMSE < 10%

4 RESULTADOS E DISCUSSÃO

4.1 Modelo Estatístico

4.1.1 Obtenção e validação das equações de H-S para as 11 cidades

A figura 5 mostra as correlações da irradiação global diária (H_G) e diferença de temperaturas, máxima e mínima, ($\Delta T = T_{\max} - T_{\min}$) com o modelo H-S bem como as curvas de regressão obtidas a partir das correlações de H_G/H_0 e $(\Delta T)^{0.5}$ para o modelo H-S, tabela 1 das onze (11) cidades do Estado de São Paulo.

Figura 5- Correlações da irradiação global diária.

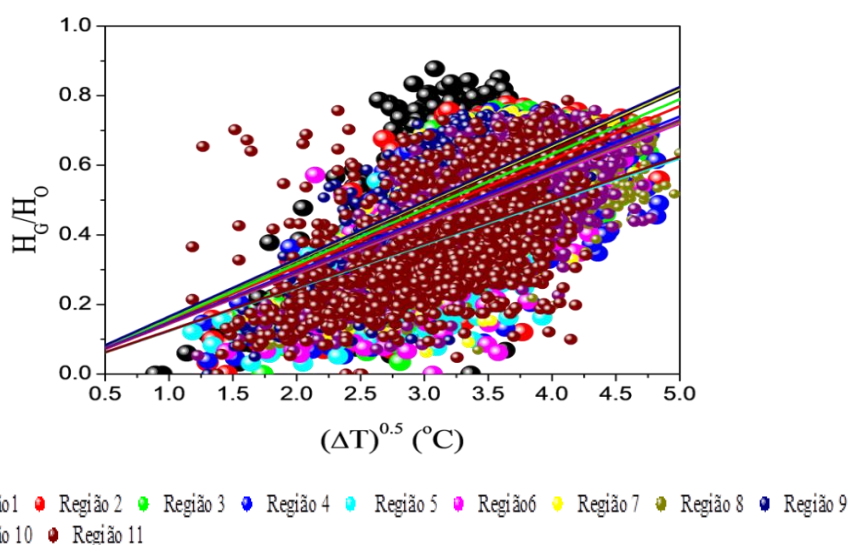
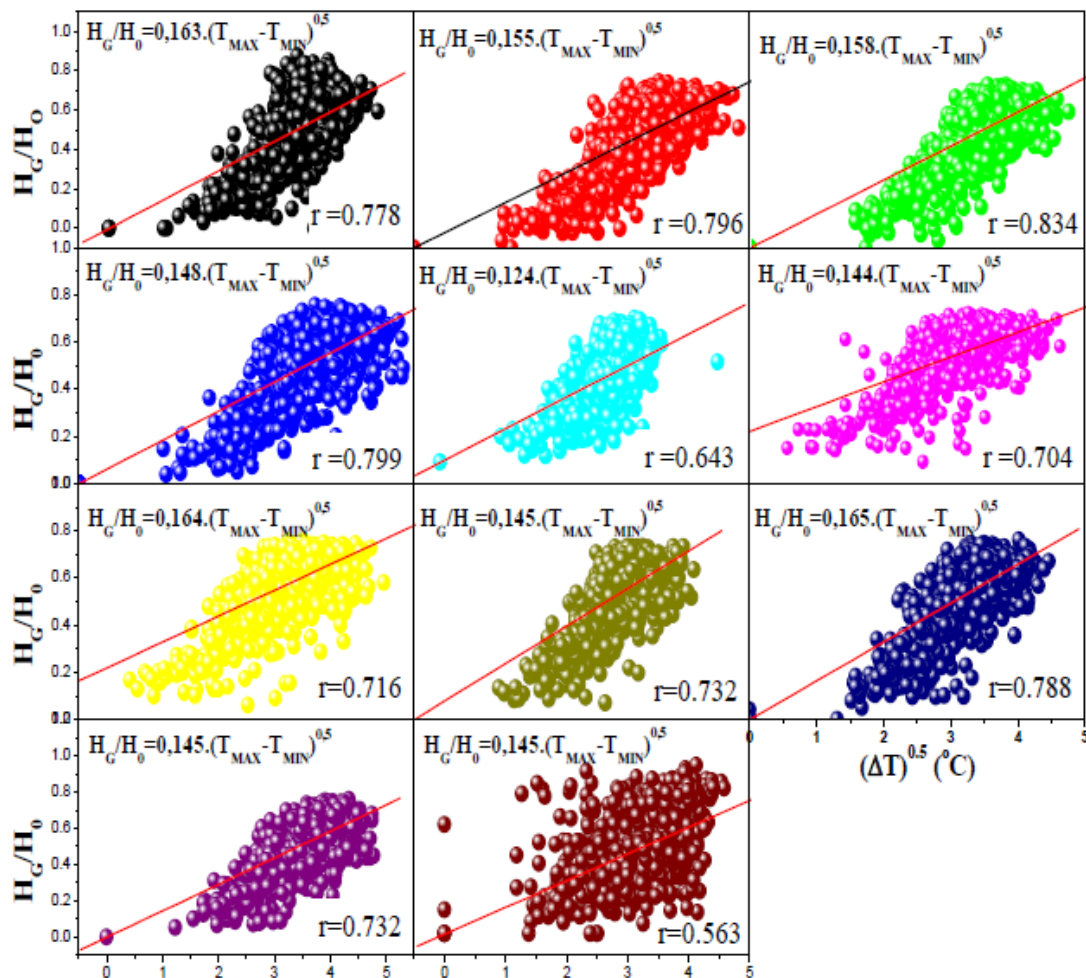


Figura5. Correlações da transmissividade da irradiação global diária (H_G/H_0) e diferença de temperaturas, máxima e mínima, $\Delta T^{0.5} = (T_{\max} - T_{\min})^{0.5}$ e curvas de regressão em função das correlações de H_G e ΔT para o modelo H-S, as onze(11) cidades do Estado de São Paulo.

A distribuição dos pontos experimentais mostra elevado espalhamento, onde constata-se que há uma larga variação de H_G/H_0 para mesmo valor de $\Delta T^{0.5} = (T_{\max} - T_{\min})^{0.5}$ em todo intervalo de variação de $\Delta T^{0.5}$. Para um mesmo local, latitude e altitude constantes, existem muitos valores de H_G/H_0 o mesmo valor de $(\Delta T)^{0.5}$ devido às variações de clima, forçados por eventos micro e macro climático, além das variações sazonais ou naturais. A variabilidade de H_G/H_0 é devido a variações de ΔT , as quais dependem do clima, latitude e altitude, de cada local. É conhecido na literatura, que os maiores valores de ΔT ocorrem em locais de maiores altitude e latitude, as quais apresentam maiores amplitudes térmicas,

Devido à temperatura mínima ser mais baixas ao longo do ano no inverno, determinando um aumento de ΔT . Ao contrario, os menores valores de ΔT ocorrem em locais de menores valores da altitude e latitude, nas quais a temperatura mínima é mais próxima da temperatura máxima, independente das estações do ano.

Figura 6- Curvas de regressão linear para o modelo HS das 11 cidades com seus respectivos coeficientes de determinação.



As curvas do modelo H-S são lineares do tipo $H_G/H_0 = a \cdot (T_{\max} - T_{\min})^{0.5}$ como mostrado figura 6.

A variabilidade do coeficiente angular a expressam a dependência dos fatores: clima, latitude e altitude nas estimativas de H_G/H_0 pelas equações. Na tabela 6 são apresentados os valores do coeficiente, a bem como do coeficiente de determinação das equações H-S das onze (11) cidades do Estado de São Paulo.

**Tabela 6- Regiões Geográficas, cidades, altitudes, latitudes e climas
coeficiente de correlação a de H-S.**

Regiões Geográficas	Altitude	Clima	a	r ²
1.São Paulo(São Paulo)	792,06	Subtropical úmido	0,163	0,608
2.Sorocaba (Sorocaba)	609	Subtropical	0,155	0,624
3.Bauru (Lins)	459	Tropical de altitude	0,158	0,672
4.Marília (Ourinhos)	448	Subtropical	0,148	0,640
5.Presidente Prudente (Presidente Prudente)	435,55	Tropical	0,124	0,410
6.Araçatuba (Ilha Solteira)	337	Tropical semi-úmido	0,144	0,490
7.São Jose do Rio Preto (Jales)	457	Tropical	0,164	0,504
8.Ribeirão Preto (Barreto)	533	Tropical com estação seca	0,145	0,533
9.Araraquara (São Carlos)	863	Tropical de altitude	0,165	0,624
10.Campinas (Itapira)	633	Tropical de altitude	0,146	0,533
11.São Jose dos Campos (Campos do Jordão)	1642	Oceânico	0,124	0,314

Os maiores valores do coeficiente angular a foi obtido para a cidade de São Carlos a = 0,165 (altitude 863m e clima tropical de altitude) e Jales a = 0,164 (altitude 457m e clima tropical) em diferentes condições de latitude, altitude e clima. Já, o menor valor do coeficiente a = 0,124 foi obtido em campos de Jordão (altitude é de 1642m, e clima oceânico) e Presidente Prudente (altitude 436m e clima tropical) também em diferentes condições de latitude, altitude e clima. Era esperado que os valores de (a) fossem diferentes entre as cidades nos dois casos, principalmente devido à diferença da altitude. Outro caso comparativo interessante se deu nos valores de a das cidades de Itapira (a = 0,146), Barretos (a = 0,145), Ilha Solteira (a = 0,144) e Ourinhos (a = 0,148), os quais foram aproximadamente iguais, com largas diferenças entre latitude, altitude e clima, entre cidades. É possível haver uma compensação entre os efeitos da latitude, longitude e clima, nas variações de $\Delta T^{0,5}$.

Os valores a obtidos para as equações de estimativas $H_G/H_0 = a \Delta T^{0,5}$ variando de a = 0,124 até a = 0,165 estão em concordância com resultados obtidos em outros locais do globo terrestre, conforme mostra os valores de a na tabela 7.

Tabela7- Localidades Altitudes Clima e Coeficiente de H_G.

Autores	Localidades	Latitude	Altitude	Clima e solo	Coeficiente a	r²
Almorox et al. (2013)	Córdoba (Argentina)	30° 44' S		Subtropical úmido	0,146	
Almorox et al. (2011)	Madri (Espanha)	40° 04' - 40° 24' N	477 - 604	Mediterrâneo continental temperado região semi-árida	0,146 - 0,161	
Almorox (2009)	Aranjuez (Espanha)	40° 04' N			0,146	0,87
Estévez et al. (2012)	65 (Espanha)	36° 17' - 38° 15' N	4 - 1212	Semiárido	0,160 - 190	
Gois et al. (2016)	Seropédica (Brasil)	22° 44' S	26	Tropical chuvoso	0,16 - 0,19 (predeterminado)	0,93
Grillone et al. (2012)	33 Sicília (Itália)	36,96° - 38,48° N	40 - 1840	Mediterrâneo	0,146 a 0,215	
Lêdo et al. (2012)	Barbalha	7° 18' S	409	Tropical quente semiárido	0,16 (predeterminado)	0,31
Liu et al. (2009)	15 (China)	34° 18' - 45° 45' N	3 a 2295		0,141 - 0,166	
Mehdizadeh et al. (2016)	Kerman (Irã)	30° 15'	1753,8	Arido	0,154	0,809
Nogueira et al. (2012)	Jaguaruana	4° 47' S	12		0,16 (predeterminado)	0,34
Rahimikhoob (2010)	Ahwaz (Irã)	31° 20'	22,5	Semi-árido	0,156	0,84
Sanusi et al. (2013)	Sokoto (Nigéria)	13° 03'			0,11	0,146
Sharifi et al. (2016)	2 (Irã)	35° 58' - 39° 46' N	1313 - 1364	Árido e semi-árido	0,162 e 0,165	0,781 e 0,809
Sharifi et al. (2016)	4 (EUA)	40° - 43° N	378 - 1208	Semi-árido	0,136 a 0,147	0,709 a 0,802
Ugwu e Ugwuanyi (2011)	Abuja (Nigéria)	8° 15' - 9° 12' N				0,790

Fonte: baseado no UNEP 1993

A variabilidade do valor de **a** em cada cidade, em relação ao valor médio total de **a** do estado, foi calculada pela equação (2):

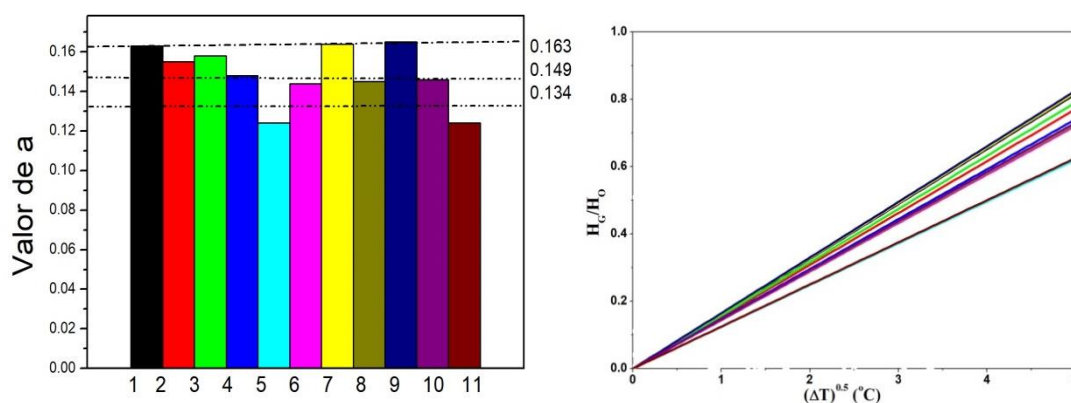
$$D(\%) = 100 * \left(\frac{a_a - a_c}{a_a} \right) \dots\dots\dots (2)$$

E os resultados estão apresentados na tabela 8. O valor de D(%) expressa a capacidade do valor de o médio subestimar ou superestimar o valor de a de cada cidade, como podem ser observado na figura 7.

Tabela 8- Cidades, Altitude, Coeficiente de H-S, Desvio Padrão e Porcentagem.

Regiões Geográficas	Altitude	a	Desvio	%
1).São Paulo(São Paulo)	792,06	0,163	0,096	96,0
2).Sorocaba (Sorocaba)	609	0,155	0,042	42,2
3).Bauru (Lins)	459	0,158	0,062	62,3
4).Marília (Ourinhos)	448	0,148	0,005	0,5
5).Presidente Prudente (Presidente Prudente)	435,55	0,124	0,166	16,6
6).Araçatuba (Ilha Solteira)	337	0,144	0,032	31,8
7).São Jose do Rio Preto (Jales)	457	0,164	0,103	10,3
8).Ribeirão Preto (Barreto)	533	0,145	0,025	25,1
9).Araraquara (São Carlos)	863	0,165	0,109	10,9
10).Campinas (Itapira)	633	0,146	0,018	18,3
11).São Jose dos Campos (Campos do Jordão)	1642	0,124	0,166	16,6

Figura 7- Gráficos de coeficiente de H-S das 11 cidades.

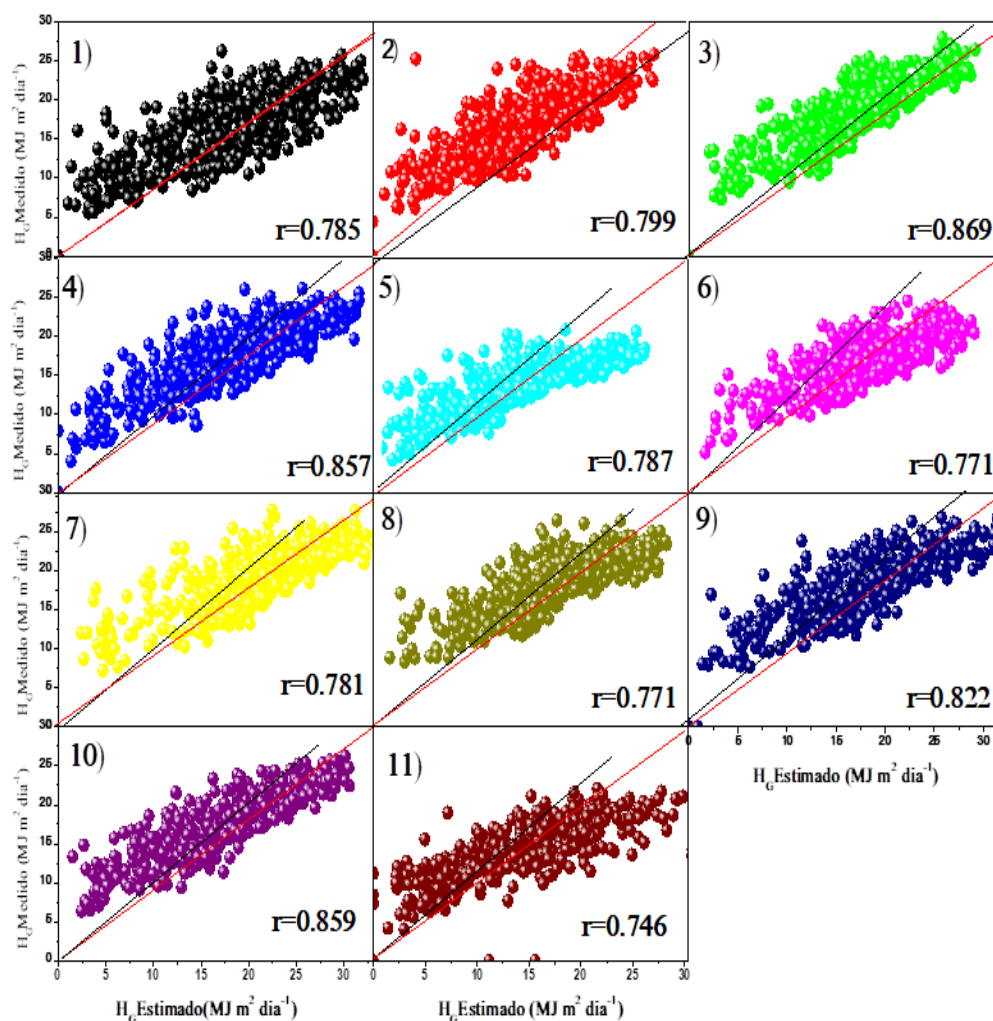


Os valores de D(%) obtidos mostram que o valor **a (médio) = 0,149** anual superestima o valor **a** nas cidades de: Campos de Jordão ($a = 0,124$) em 16,6%; Presidente Prudente ($a = 0,124$) em 16,6%; Ilha Solteira ($a = 0,144$)

em 3,2%; Barretos ($a = 0,145$) em 2,5% Itapira ($a = 0,146$) em 1,8% e ourinhos ($a = 0,148$) em 0,5%. Similarmente, o valor **a (médio) = 0,149** anual subestima o valor **a** nas cidades de: São Carlos ($a = 0,165$) em 10,9%; Jales ($a = 0,164$) em 10,3%; São Paulo ($a = 0,163$) em 9,6%; Lins ($a = 0,158$) em 6,2% e Sorocaba ($a = 0,155$) em 4,2%. Os resultados variando de 17% a 0,5%, entre as regiões do estado de São Paulo, mostram diferenças significativas nos valores de **a** em relação ao valor **a** médio, nas duas condições de estimativas.

As Fig.8 (1-11) mostram as correlações obtidas nas validações, comparação entre estimativa de H_g por meio das Eq. e a medida H_g selecionada para validação.

Figura 8- Comparação entre os valores estimados de H_g pelas equações 1-11(de cada cidade) e medidas de H_g com seus respectivos coeficiente de correlação.



As correlações entre os valores estimados e medidos estão em concordância com as curvas ideais de 45° ($r = 1$) nos ajustes, e são lineares nos intervalos de variação das medidas de H_G/H_0 . Os valores do coeficiente de correlação (r), variaram entre as 11 cidades de $r = 0,77$ a $r = 0,87$, indicam que estatisticamente as estimativas das equações 1-11 estão correlacionadas com as medidas, dentro da ordem de grandeza da precisão da equação de Angstrom-Prescott, considerada a equação de referencia nas estimativas da radiação global.

Valores dos indicadores estatísticos MBE e RMSE obtidos na validação das equações são mostradas na tabela 9.

Tabela 9- Valores dos indicadores estatísticos MBE,e RMSE obtidos na validação das equações 1-11.

Regiões Geográficas	Cidade	MBE (MJ m ² dia ⁻¹)	rMBE (%)	RMSE (MJ m ² dia ⁻¹)	rRMSE (%)
1.São Paulo	São Paulo	0,1666	1,07	4,284	27,391
2.Sorocaba	Sorocaba	0,1954	1,19	3,938	23,913
3.Bauru	Lins	0,2406	1,41	3,542	20,721
4.Marília	Ourinhos	0,2588	1,61	3,826	23,843
5.Presidente Prudente	Presidente Prudente	0,2290	1,72	3,412	25,664
6.Araçatuba	Ilha Solteira	0,1116	0,67	3,369	20,348
7.São Jose do Rio Preto	Jales	0,1074	0,57	3,600	19,177
8.Ribeirão Preto	Barreto	0,1527	0,86	3,678	20,784
9.Araraquara	São Carlos	0,1809	1,03	3,727	21,171
10.Campinas	Itapira	0,3710	2,18	3,753	22,077
11.São Jose dos Campos	Campos do Jordão	0,3023	2,23	4,508	33,249

Os valores obtidos para $rMBE$ mostram que a equação 1 a11 superestima a medida com percentual inferior a 2,5%. Os valores de $rMBE$ são inferiores as incertezas das medidas das irradiâncias global e tempertura que são da ordem de 10,0%. Os valores obtidos para $rRMSE$ mostram que o espalhamento ou desvio do erro médio das Equações 1-11 inferiores a 33,0% é considerado regular estatisticamente (Heinemann e Maia, 2012; Jamieson e Wilson, 1991; Li, e Liu, 2013) e da mesma ordem de grandeza aos resultados obtidos com modelos estatísticos com outras radiações ou frações. (Da Silva et al., 2017).

4.1.2 Obtenção e Validação da Equação H-S Geral ou Espacial

A figura 9 mostra a correlações da irradiação global diária (H_G) e diferença de temperaturas, máxima e mínima: $\Delta T^{0,5} = (T_{\max} - T_{\min})^{0,5}$ espacial com o modelo H-S para uma base de dados com as onze(11) cidades juntas do Estado de São Paulo.

Figura 9- Correlações da irradiação global diária (H_G).

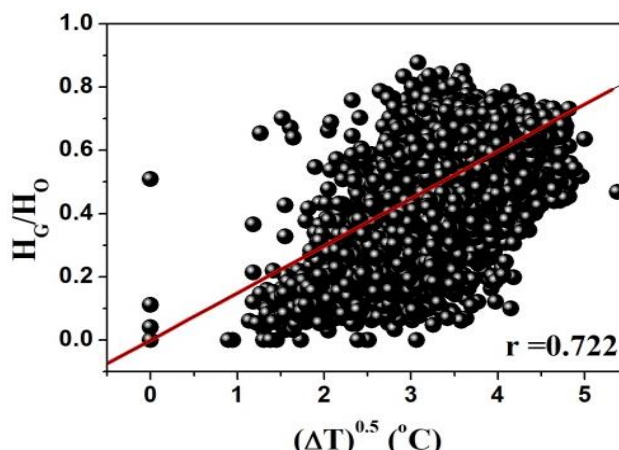


Figura 9 Correlação da irradiação global diária (H_G) e diferença de temperaturas, máxima e mínima $\Delta T = (T_{\max} - T_{\min})^{0,5}$ com o modelo H-S para base de dados total. Curva de regressão linear com seus respectivos coeficientes de determinação r .

Similarmente a descrição da secção 4.1.1, a distribuição dos pontos experimentais para a base de dados total mostra também elevado espalhamento, com larga variação de H_G/H_0 para mesmos valores de $\Delta T^{0,5} = (T_{\max} - T_{\min})^{0,5}$ em todo intervalo de variação de $\Delta T^{0,5}$. O valor de $r = 0,722$ para a equação geral é da mesma ordem de grandeza as correlações para algumas cidades e inferior a outras.

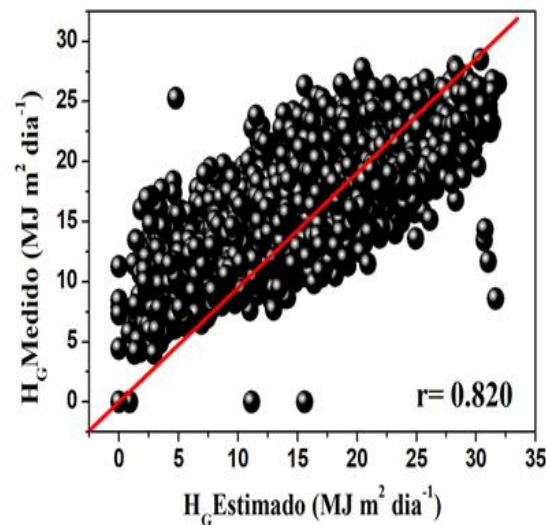
A equação obtida foi:

$$H_G/H_0 = 0,149 \Delta T^{0,5} \dots\dots\dots(3)$$

com $r^2 = 0,596$

Similarmente, o valor de $r^2 = 0,596$ para a equação geral é da mesma ordem de grandeza aos valores de r^2 das equações obtidas para algumas cidades e inferior a outras.

Figura 10- Comparação entre os valores estimados de H_G pela equação e medidas de H_G com seu respectivo coeficiente de correlação.



Os valores dos indicadores estatísticos rMBE e rRMSE obtidos na validação das equações foram: rMBE = 1,3% e rRMSE = 23,3%. Os valores obtidos para a equação geral estão dentro do intervalo de variação dos valores obtidos para as 11 cidades.

4.2 Treinamento e Validação das Técnicas RNA e MVS

4.2.1 Treinamento e Validação das Técnicas MLP e MVS

As tabelas mostram o coeficiente de parametrização usado no treinamento de redes: tipo de função, número de camadas e épocas e os indicadores estatísticos de rRMSE para a combinação 1 e 5, para as onze cidades do Estado de São Paulo para as duas redes (MLP e MVS). Nas tabelas são apresentados o desempenho de treinamento das duas redes para as Combinações de 1-5, e mostram que o pior desempenho ocorreu para a combinação 1 (menor número de variáveis), e o melhor desempenho ocorreu para a combinação 5. Para facilitar a discussão apresentaremos as tabelas 10 que mostram o pior e o melhor resultado e a tabela da combinação (1) que representa a mesma entrada de variáveis do modelo estatístico.

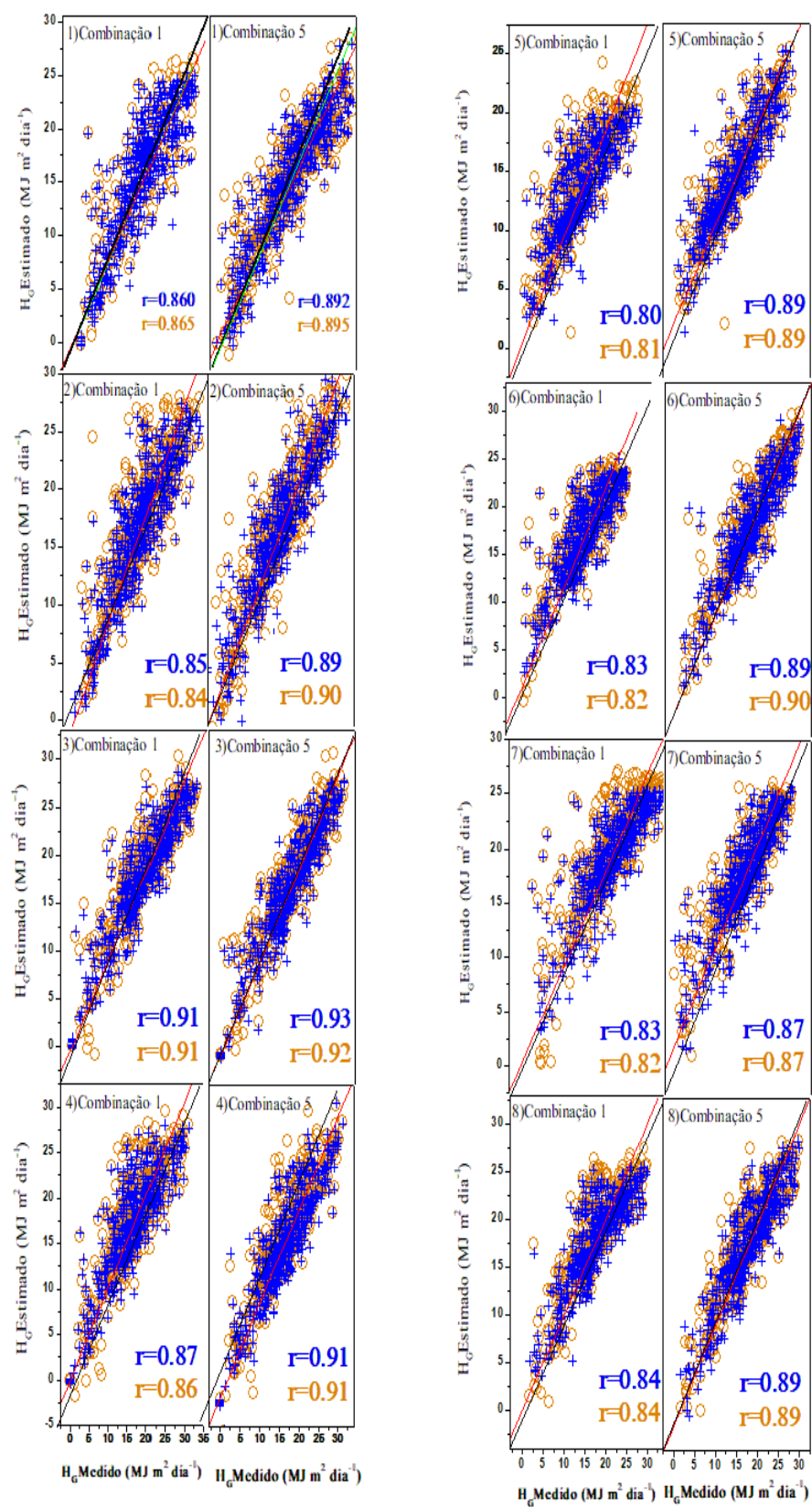
Tabela 10- Coeficientes de parametrização das combinações 1 e 5 para as onze cidades do Estado de São Paulo.

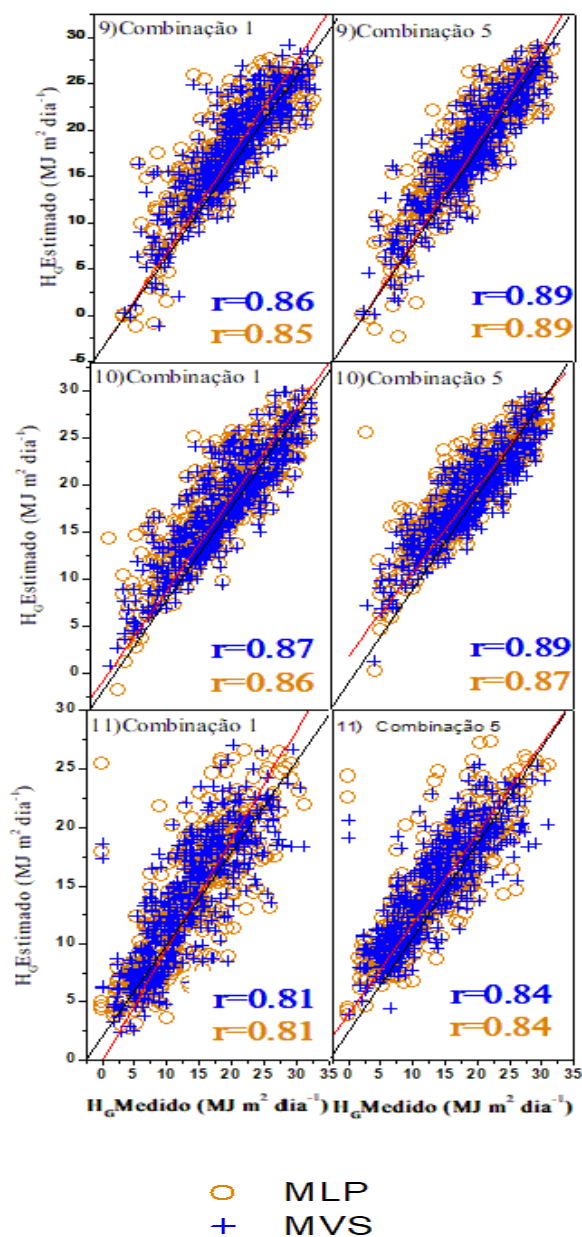
TREINAMENTO								
REDE MLP				REDE MVS				
Combinação1	função de	camada 1	Épocas	rRMSE (%)	coeficiente de parametrização			
	ativação	de ajustamento			C	γ	ε	rRMSE (%)
São Paulo	Tang. Hiperb.	n1-5	21	22.98	100	0.3	0.001	27.81
Sorocaba	Tang. Hiperb.	n1-5	12	20.99	100	0.3	0.001	23.42
Lins	Tang. Hiperb.	n1-5	19	15.97	100	0.3	0.001	21.20
Ourinhos	Tang. Hiperb.	n1-5	7	20.90	100	0.3	0.001	26.69
Presidente P.	Tang. Hiperb.	n1-5	26	23.43	100	0.3	0.001	24.84
Ilha Solteira	Tang. Hiperb.	n1-5	9	19.19	100	0.3	0.001	19.80
Jales	Tang. Hiperb.	n1-5	12	16.91	100	0.3	0.001	19.20
Barreto	Tang. Hiperb.	n1-5	16	17.37	100	0.3	0.001	17.37
São Carlos	Tang. Hiperb.	n1-5	47	18.37	100	0.3	0.001	23.20
Itaipira	Tang. Hiperb.	n1-5	14	18.07	100	0.3	0.001	21.75
Campos de Jordão	Tang. Hiperb.	n1-5	17	27.64	100	0.3	0.001	35.55
Combinação 5								
São Paulo	Tang. Hiperb.	n1-5	21	20.67	100	0.3	0.001	21.19
Sorocaba	Tang. Hiperb.	n1-5	12	17.37	100	0.3	0.001	16.73
Lins	Tang. Hiperb.	n1-5	19	14.69	100	0.3	0.001	15.52
Ourinhos	Tang. Hiperb.	n1-5	7	17.99	100	0.3	0.001	19.25
Presidente P.	Tang. Hiperb.	n1-5	26	19.54	100	0.3	0.001	19.66
Ilha Solteira	Tang. Hiperb.	n1-5	9	14.83	100	0.3	0.001	14.71
Jales	Tang. Hiperb.	n1-5	12	15.37	100	0.3	0.001	15.53
Barreto	Tang. Hiperb.	n1-5	16	14.65	100	0.3	0.001	14.65
São Carlos	Tang. Hiperb.	n1-5	47	16.18	100	0.3	0.001	17.77
Itaipira	Tang. Hiperb.	n1-5	14	20.09	100	0.3	0.001	18.11
Campos de Jordão	Tang. Hiperb.	n1-5	17	27.24	100	0.3	0.001	28.56

A comparação do valor de rMBE ou desempenho entre as 2 tabelas mostra que quanto menor o numero de variável de entrada, maior o valor de rRMSE no treinamento sem que haja um valor sistemático superior ou inferior na 5 combinações, para as duas a redes : MLP e MVS.

Para a rede MLP, valor de rRMSE variou entre as 11 onze cidades para 5 combinações ondes o maior valor ocorreu para a combinação 1 com valor variando entre 15,97% (Lins) a 27,64% (Campo de Jordão), enquanto que o menor valor ocorreu entre 14,65% (Barreto) a 27,24% (Campos de Jordão). Para rede MVS, valor de rRMSE variou entre as 11 onze cidades para 5 combinações ,onde o maior valor ocorreu para a combinação 1 com valor variando entre 17,37% (Barreto) a 35,55% (Campos de Jordão), enquanto que os menores valores ocorreu entre 14,65%(Barreto) a 28,56% (Campo de Jordao).

Figura 11- Diagramas de dispersão entre os valores estimados de H_G , pelos modelos gerados de MLP e MVS, para as 11 cidades do estado de São Paulo.





As entradas das variáveis das combinações 1 a 5 para as duas redes MLP e MVS elevaram o valor do coeficiente de correlação onde, os valores de r obtidos com rede MVS foram superiores aos valores da rede MLP, para as 11 cidades.

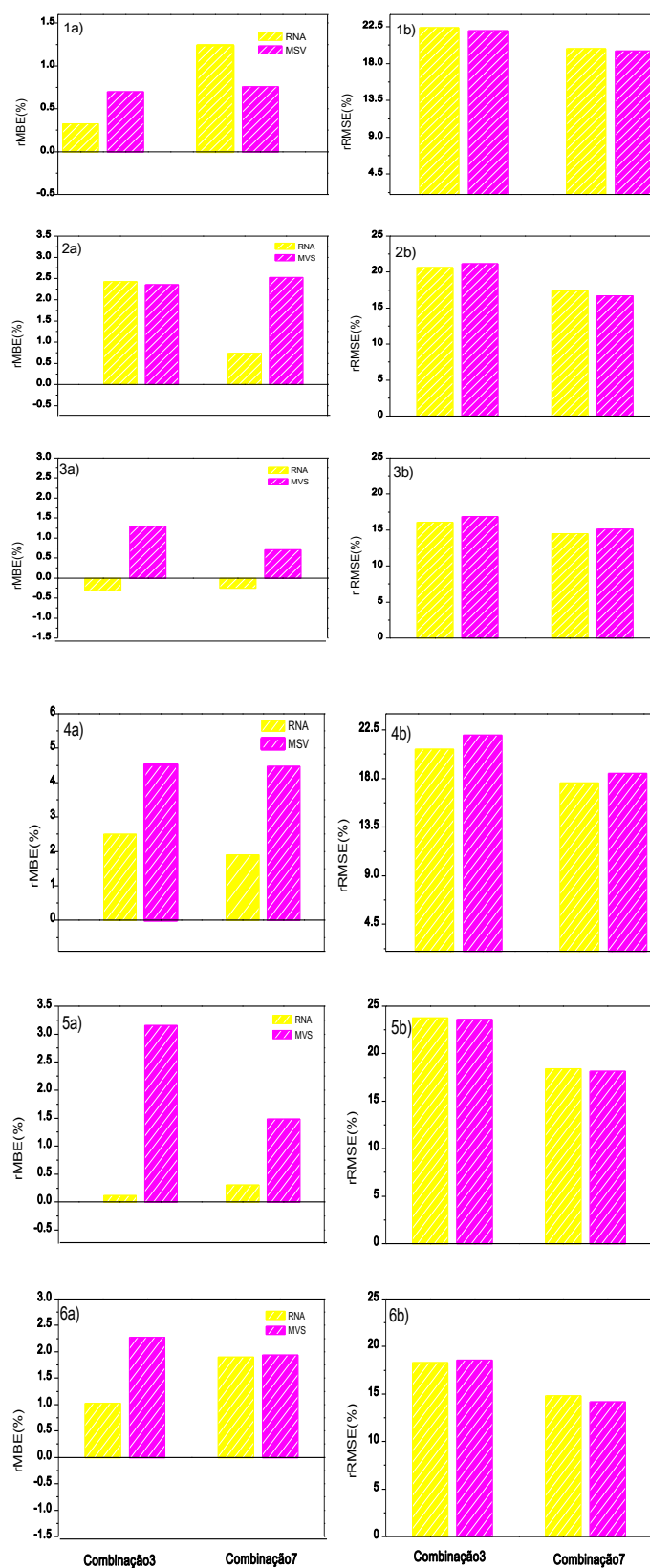
Os resultados encontrados são similares aos na literatura por: (Bocco et al., 2006; Chen et al., 2004; Elizondo, 1994; Landeras et al., 2012; Linares, 2011). Em especial, a combinação 1, que trata as mesmas variáveis do modelo clássico ou estatístico.

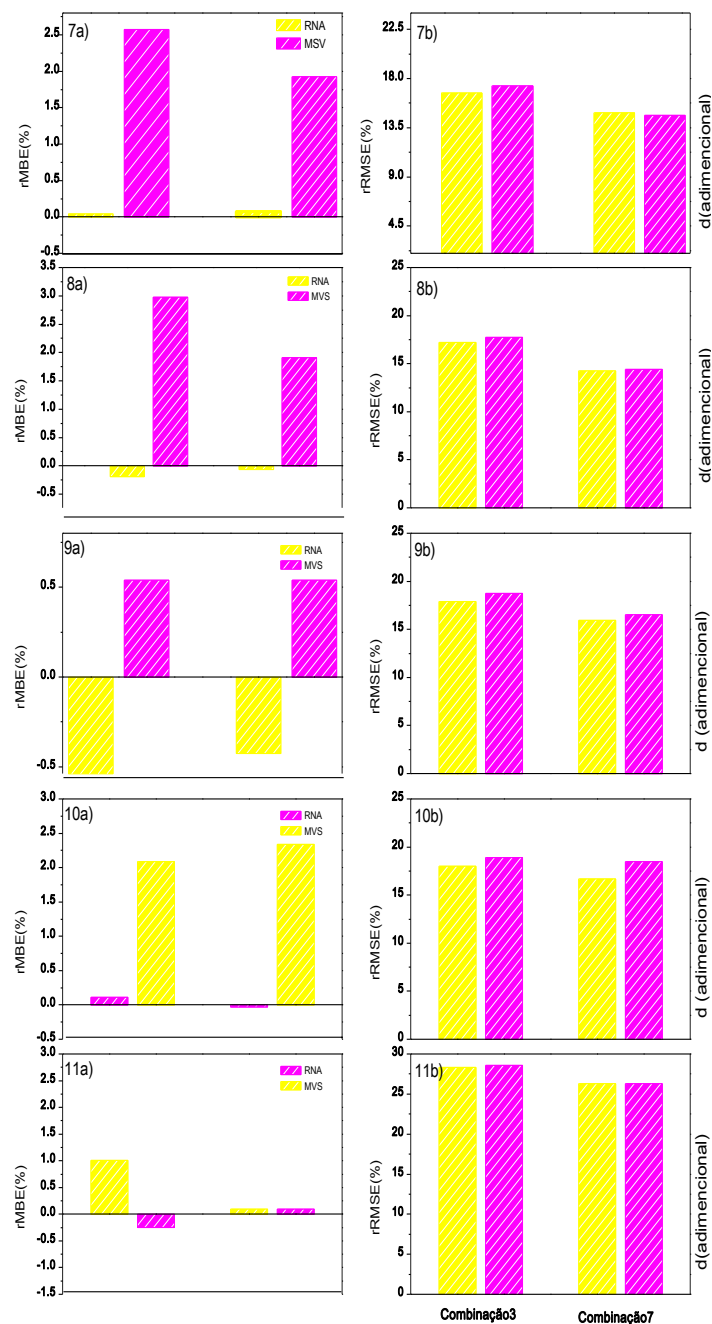
A combinação das variáveis H_o , $T_{máx}$ e $T_{mín}$, nos modelos MLP1 e MVS1, representam a junção de todas as variáveis das equações de Angstrom-Prescott (A-P), Hargreaves-Samani. Para cidade de São Paulo, com as redes $RNA_{(MLP)1}$ e MVS1, os coeficientes de correlação de $r = 0,860$ e $r = 0,865$ respectivamente, mostram elevadas correlações, mais elevadas que o modelo clássico.

Os modelos de rede elevaram o valor de r em relação aos modelos estatísticos da ordem de 0,6–0,7 para valores de 0,8 a 0,9. Esses resultados foram considerados importantes, porque mostrou que a entrada de novas variáveis permite o aumento de desempenho dos modelos de rede. Para a combinação 5, com o maior número de variáveis (Figura 11) os valores de r foram os mais elevados, com coeficientes de correlação $r = 0,892$ e $r = 0,895$ pelas MLP5 e MVS5 respectivamente. A introdução da umidade relativa (%) na combinação 4 manteve os valores de r , aproximadamente iguais na condição de validação, aumentando o valor de $r = 0,885$ para r .

A figura 12 (1-11) mostra diagrama dos valores de r_{MBE} e r_{RMSE} entre os valores estimados de H_G , e pelos modelos gerados pelas MLP e MVS, respectivamente, para as combinações 1 e 5 para as 11 cidades do estado de São Paulo.

Figura 12- Indicativos estatísticos rMBE, rRMSE na validação para a combinação 1 e 5 .





Para a combinação 1, onde os valores de entrada nas redes são iguais as variáveis do modelo estatístico, os valores de rRMSE, ou desempenho do modelo, não foram iguais em todas as cidades como mostra a tabela 11: para MLP, o valor de rRMSE variou de 16,11 em Lins a 28,32 Campos de Jordão, e para a rede MVS o valor de rRMSE, variou de 16,88 em Lins a 28,62 Campos de Jordão. Similarmente, os valores de rMBE, inferiores a 0,5% também não foram iguais

para as 11 cidades: para a rede MLP ocorreu um numero de superestimava maior (8) que a subestimativa(3) do modelo, enquanto que para a rede MVS, inferiores a 4,53%, ocorreu superestimava do modelo nas 11 cidades. A comparação dos valores de rRMSE entre as duas redes MLP e MVS mostra que o desempenho da estimativa de H_G para essa combinação não haver uma supremacia, de uma rede sobre outra entre as 11 cidades. As redes MVS e MLP apresentaram melhor desempenho nas cidades de: Lins, Jales, Barretos.

Tabela 11 - Combinação 1 para os modelos da redes de MLP e MVS.

Combinação 1	MLP		MVS	
	rMBE	rRMSE(%)	rMBE	RMSE(%)
São Paulo	0,05	22,40	0,70	22,04
Sorocaba	0,40	20,66	2,36	21,18
Lins	-0,05	16,11	1,29	16,88
Ourinhos	0,40	20,74	4,53	22,03
Presidente P.	0,02	23,79	3,16	23,62
Ilha Solteira	0,17	18,31	2,28	18,59
Jales	0,01	16,69	2,58	17,34
Barreto	0,03	17,25	2,98	17,74
São Carlos	0,10	17,93	0,54	18,75
Itapira	0,02	18,01	2,09	18,94
Campos de Jordão	-0,13	28,32	0,24	28,62

Para a combinação 5, onde os valores de entrada nas redes é máximo (5), o desempenho do modelo que mostra na tabela 12, não foram iguais em todas as cidades: para MLP, o valor de rRMSE variou de 14,26 em Barreto a 26,35 Campos de Jordão, e para a rede MVS o valor de rRMSE, variou de 14,21 em Ilha Solteira a 26,35 Campos de Jordão. Similarmente, os valores de rMBE, inferiores a 2% também não foram iguais para as 11 cidades: para a rede MLP ocorreu um numero de superestimava maior (7) que a subestimativa(4) do modelo, enquanto que para a rede MVS, inferiores a 4,48%, ocorreu superestimava do modelo nas 11 cidades.

Tabela 12- Combinação 5 para os modelos de redes de MLP e MVS.

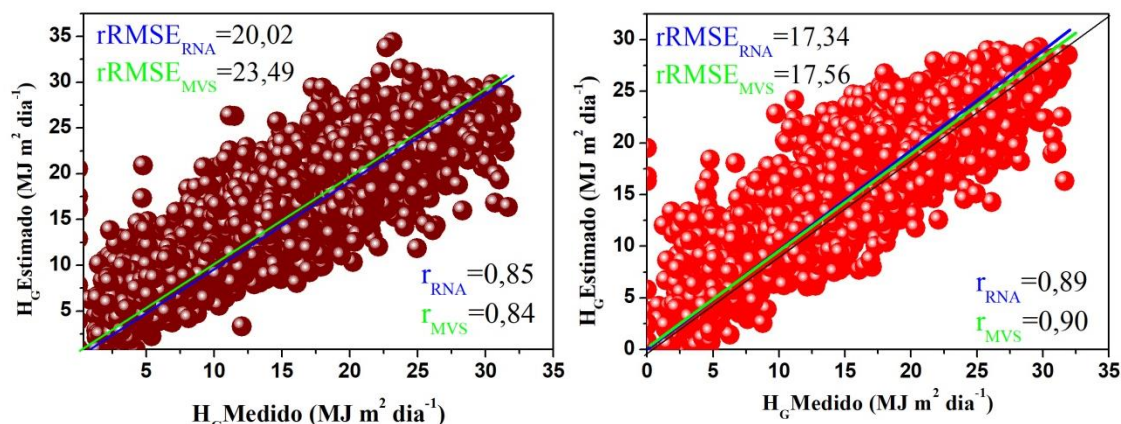
Combinação 5	MLP		MVS	
	rMBE(%)	rRMSE(%)	rMBE(%)	RMSE(%)
São Paulo	0,19	19,83	0,76	19,56
Sorocaba	0,12	17,41	2,53	16,73
Lins	-0,04	14,49	0,71	15,17
Ourinhos	0,30	17,61	4,48	18,49
Presidente P.	0,04	18,41	1,49	18,18
Ilha Solteira.	0,03	14,83	1,94	14,21
Jales	0,07	14,89	1,93	14,65
Barreto	0,01	14,26	1,91	14,45
São Carlos	0,08	15,96	0,54	16,54
Itapira	0,03	16,71	2,34	18,49
Campos de Jordão	0,10	26,35	0,47	26,35

A comparação dos valores de RMSE entre as duas redes MLP e MVS mostra que o desempenho da estimativa de H_G para combinação 5 nas 11 cidades também não haver uma superioridade, de uma rede sobre outra entre as 11 cidades. A rede MVS foi superior nas cidades de São Paulo, Sorocaba, Presidente Prudente, Ilha Solteira e Jales, enquanto que e MLP foi superior nas cidades de Lins, Ourinhos Barretos, São Carlos, Itapira em Campos de Jordão as duas redes tiveram desempenho em estimar H_G iguais.

4.2.2. Desempenho do modelo de redes Total (r e rRMSE)

A Figura 13 mostra a correlação obtida da validação, comparação entre estimativa de H_G por meio das redes MLP e MVS e a medida H_G selecionada para validação na combinação 1 e 5.

Figura 13- Desempenho do modelo de redes Total da combinação 1 e combinação 5.



Para a combinação 1, os valores do coeficiente de correlação $r_{RNA(MLP)} = 0,850$ e $r_{MVS} = 0,841$ obtidos por meio das redes MLP e MVS foram superiores ao valor $r = 0,820$ do modelo estatístico, demonstrando que nas mesmas condições de entrada das variáveis, o desempenho das redes em estimar H_G é superior ao modelo estatístico. Os valores do coeficiente de correlação $r_{RNA(MLP)} = 0,889$ e $r_{MVS} = 0,901$ obtidos por meio das redes MLP e MVS na combinação 5 foram superiores ao valor $r = 0,820$ com melhor desempenho das redes em relação ao do modelo estatístico.

Os valores dos indicadores estatísticos RMSE obtidos na validação para a combinação 1 das redes MLP e MVS: $rRMSE_{RNA(MLP)} = 20,02\%$ e $rRMSE_{MVS} = 23,5\%$ foram da mesma ordem de grandeza do resultado obtido para o modelo estatístico $RMSE = 23,3\%$. Portanto, desempenho entre os modelos de redes e estatístico é aproximadamente iguais. Para a combinação 5, os valores dos indicadores estatísticos $rRMSE$ obtidos das redes MLP e MVS: $rRMSE_{RNA(MLP)} = 17,4\%$ e $rRMSE_{MVS} = 17,5\%$ foram inferiores do resultado obtido com o modelo estatístico $RMSE = 23,3\%$. Portanto, melhor desempenho para o modelo de redes, os quais foram praticamente iguais.

4.2.3 Comparação do desempenho do modelo MVS e RNA de com resultados da literatura

A tabela 13 mostra a comparação dos indicadores estatísticos $rMBE$, MBE , $rRMSE$, $RMSE$, em unidades de energia por dia e em porcentagem, r ou R^2 obtidos nas condições de validações, dos modelos MVS e RNA com as onze regiões de São Paulo em várias localidades do mundo.

Tabela 13- Comparação do desempenho dos modelos com resultados da literatura.

Autores	Localidades	Modelos	MBE	rMBE	RMSE	rMSE	r	R2
	N'	Redes neurais	(MJ m-2 dia-1)	%	(MJ m-2 dia-1)	%		
Bocco et al. (2010)	Salta (Argentina)	RNA			1,66 a 2,97			0,73 a 0,92
Bocco et al. (2006)	Córdoba (Argentina)	RNA			3,14 a 3,88		0,86 a 0,92	0,96 a 0,99
Bosch et al. (2008)	Granada (Espanha)	RNA		-1,2 a 2,1		5,0 a 7,5		0,98
Elizondo et al. (1994)	3 (Estados Unidos)	RNA			2,92 a 3,64			
Fortin (2008)	montreal (Canada)	RNA			3.83 a 5.45			
Landeras et al. (2012)	País Basco (Espanha)	RNA			2,93			
Linarez-Rodríguez et al. (2011)	Andaluzia (Espanha)	RNA	-0,470		2,88	16,4	0,94	
Lyra et al. (2015)	Alagoas(Brasil)	RNA			1.97	10.1		0,75
Mejdoul et al. (2013)	8(Marrocos)	RNA			1,23 a 1,5		0,98	
Waewasak et al. (2014)	Bangkok (Tailândia)	RNA	-0,020 - 0,003		0,003 a 0,392			
Urraca et al. (2015)	4 (Espanha)	RNA			2.11 - 2.9			0.90 - 0.93
Modelo do trabalho	11 (São Paulo)	RNA	0.04 a 0.19	0.45a 1,90	2,4 a 3,5	14,26 a 26,35	0,83 a 0,93	
Chen et al. (2015)	32 (China)	MVS			0,840 a 2,660			
Chen et al. (2013)	3 (China)	MVS			1,79 a 2,38	13,4 a 17,4		
Mohammadi et al. (2015)	Bandar Abbas (Irã)	MVS			1.424 a 2.691			0.887 a 0.899
Quej et al. (2015)	4 (mexico)	MVS			2,40 a 3,66			0,48 a 0,7
Quej et al. (2017)	Yucatán (México)	MVS			2.578 a 2.877			
Ramedani et al. (2014)	Teerã (Irã)	MVS			3,30 a 3,40			0.887 a 0.899
Modelo do trabalho	11 (São Paulo)	MVS	0,12 a 0,71	0.47 a 4.48	1,9 a 3,5	14,21 a 26,35	0,84 a 0,92	

Os resultados obtidos na validação para as duas redes das onze cidades do Estado de São Paulo também estão apresentados na tabela. Para a rede $RNA_{(MLP)}$ os resultados dos valores de $rRMSE$ ou $RMSE$, bem como r para as 11 cidades do estado de São Paulo, estão em concordância, e são similares aos resultados da grande maioria dos pesquisadores apresentados na tabela 13 Landeras et al (2012) para região Basca; Elizonto et al (1994) para os Estados Unidos; Baco et al Salta e Córdoba na (Argentina); Linares –Rodrigues et al (2011) em na região da Andaluzia(espanha) e Urraca (2015) para 4 cidades da Espanha. Ainda em relação à rede Neural ANN, os resultados obtidos foram inferiores para varias partes do mundo como mostram os trabalhos de Lyra et al (2015); Boch et al (2008); Medjdoul et al (2013) e Waewasak et al (2014). E foi superior ao trabalho de Fortin (2008). Para a rede MVS os resultados dos valores de $rRMSE$ ou $RMSE$, bem como r para as 11 cidades do Estado de São Paulo, estão em concordância, e são similares aos resultados de todos os trabalhos dos pesquisadores apresentados na tabela 13, como mostra os trabalhos: Ramedani et al (2014); Chen et al (2015); Chen et al (2013); Quej et al(2015) ; Quej et al(2017) e Mohammadi et al (2015).

4.3. Comparação do desempenho entre os modelos estatístico e as redes MVS e MLP

A tabela 14 mostra a comparação dos indicadores estatísticos $rRMSE$, em percentagem, e também os valores de r obtidos nas validações, entre medida e estimativa dos modelos estatístico de H-S e das redes AM : MLP e MVS em 11 cidades do Estado de São Paulo, bem como para todos locais juntos ou espacial, para o Estado de São Paulo.

Os valores dos indicadores estatísticos $rRMSE$, e r das redes MLP e MVS nas 11 cidades do Estado de São Paulo no geral, mostram que a rede MVS possui melhor resultado de desempenho que a rede MLP e modelo de H-S nas estimativas de H_G , respectivamente. Para a situação das 11 cidades juntas, os valores de $rRMSE$, e r obtidas com as redes MLP e MVS foram: valores de $rRMSE$ inferiores e valores r superiores, aos obtidos pelo modelo estatístico de H-S.

Tabela 14- Comparação do desempenho dos modelos.

Combinação5	Estatístico		MLP		MVS	
	rRMSE (%)	r	rRMSE(%)	r	rRMSE(%)	r
São Paulo	27,39	0,62	19,83	0,89	19,56	0,90
Sorocaba	23,91	0,64	17,41	0,89	16,73	0,91
Lins	20,72	0,87	14,49	0,93	15,17	0,92
Ourinhos	23,84	0,86	17,61	0,91	18,49	0,91
Presidente P.	25,66	0,62	18,41	0,89	18,18	0,89
Ilha Solteira	20,35	0,59	14,83	0,89	14,21	0,90
Jales	19,18	0,61	14,89	0,87	14,65	0,87
Barreto	20,78	0,77	14,26	0,89	14,45	0,89
São Carlos	21,17	0,82	15,96	0,89	16,54	0,89
Itapira	22,08	0,86	16,71	0,89	18,49	0,87
Campos de Jordão	33,25	0,75	26,35	0,84	26,35	0,84
Total das 11 cidades	23,49	0,73	17,34	0,89	17,53	0,90

5.CONCLUSÕES

Dos resultados apresentados e discutidos sobre os modelos de estimativa de H_G em partição diária; estatístico de Hargreaves-Samani, e técnicas de aprendizado de máquina MLP e MVS, propostos para o Estado de São Paulo podem concluir:

O coeficiente linear (a) da equação de Hargreaves-Samani ($H_G/H_0 = a \Delta T^{0,5}$) obtidos variando no intervalo de $a = 0,124$ a $a = 0,165$ nas 11 cidades do Estado de São Paulo, depende do clima, latitude e altitude. Os valores $a = 0,314$, $a = 0,165$ encontram-se dentro do intervalo de variação dos resultados obtidos em outras localidades do globo terrestre.

A equação de Hargreaves-Samani total (11 cidades juntas do Estado de São Paulo) obtida $H_G/H_0 = 0,149 \Delta T^{0,5}$ com coeficiente de determinação $R^2 = 0,596$ é aproximadamente igual ao valor médio de a e R^2 de cada uma das 11 cidades.

As equações de estimativa de H_G de Hargreaves-Samani obtidas por meio de regressão linear nas 11 cidades do Estado de São Paulo com coeficientes de determinação variando entre os valores $R^2 = 0,314$ a $R^2 = 0,672$ são similares de outras localidades, a exceção em regiões mais quentes ou mais frias nas quais não há correlações.

Na validação das equações de H-S os valores obtidos do coeficiente de correlação variando de $r = 0,771$ a $r = 0,859$, entre H_G estimada e H_G medida as 11 cidades, indicam que o modelo H-S pode estimar H_G com precisão similar ao modelo tradicional de Angstrom.

Os indicadores estatísticos $rMBE$, MBE , $rRMSE$, $RMSE$ também obtidos na validação: $MBE = (0,11 - 0,37 \text{ MJ m}^{-2} \text{ mês}^{-1})$ ou $(0,6 - 2,2\%)$; $RMSE = (3,3 - 4,5 \text{ MJ m}^{-2} \text{ mês}^{-1})$ ou $(19,2 - 33,2\%)$; indicam que o modelo H-S pode ser utilizado na estimativa de H_G com razoável precisão e exatidão.

Os valores obtidos na validação da equação de estimativa de H_G de Hargreaves-Samani total: coeficiente de correlação $r = 0,822$, $MBE = 1,3\%$ e $RMSE = 23.3\%$ indicam que o modelo H-S pode ser utilizado na estimativa de H_G com razoável precisão e exatidão.

As técnicas MVS e RNA na combinação 1, com as mesmas variáveis de entrada H_0 e $(\Delta T^{0,5})$ que o modelo H-S, nas duas condições: 11 cidades ou total,

mostram, que podem estimar H_G com melhores desempenho que a equação de Hargreaves-Samani tradicional.

Os valores crescentes das correlações r , entre os valores medidos e estimados de H_G pelas técnicas de aprendizado de máquina com entradas de diferentes parâmetros meteorológicos, mostram que as técnicas AM e MVS elevam o desempenho do modelo e podem ser utilizadas nas estimativas de H_G com elevadas correlações para as 11 cidades individualmente ou total.

As técnicas MVS e MLP na combinação 5, com as entradas parâmetros meteorológicos: irradiação solar global no topo da atmosfera H_0 ($\text{MJ m}^2 \text{ dia}^{-1}$); e temperatura máxima $T(^{\circ}\text{C})$ [máx]; temperatura mínima $T(^{\circ}\text{C})$ [min]; temperatura média $T(^{\circ}\text{C})$ [med]; precipitação P (mm); velocidades do vento U_2 (m s^{-1}); e umidade relativa do ar UR (%), para as 11 cidades ou total, mostram que podem estimar H_G com melhor desempenho que a equação de Hargreaves-Samani tradicional, entre todas as combinações.

No geral, a comparação entre os modelos H-S, e as redes MVS e MLP na combinação 5, quanto ao desempenho para 11 cidades e total, mostrou por meio dos valores dos indicadores estatísticos r e $r\text{RMSE}$ obtidos da validação que: os modelos de rede MVS e MLP apresentaram melhores resultados que modelo H-S, e entre os modelos de rede MVS apresentou melhor resultado que o modelo de rede MLP.

Referências

- Almorox, J., Hontoria, C., & Benito, M. (2011). Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain). *Applied Energy*. <https://doi.org/10.1016/j.apenergy.2010.11.003>.
- Alam S; Kaushik SC; Garg SN. Assessment of diffuse solar energy under general sky condition using artificial neural network. *Applied Energy* 2009; 86: 554–64.
- Assouline, D., Mohajeri, N., & Scartezzini, J. L. (2017). Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Solar Energy*. <https://doi.org/10.1016/j.solener.2016.11.045>.
- Bechini, L., Ducco, G., Donatelli, M., & Stein, A. (2000). Modelling, interpolation and stochastic simulation in space and time of global solar radiation. *Agriculture, Ecosystems and Environment*. [https://doi.org/10.1016/S0167-8809\(00\)00170-5](https://doi.org/10.1016/S0167-8809(00)00170-5).
- Bocco, M., Ovando, G., & Sayago, S. (2006). Development and evaluation of neural network models to estimate daily solar radiation at Córdoba, Argentina. *Pesquisa Agropecuaria Brasileira*. <https://doi.org/10.1590/S0100-204X2006000200001>.
- Bojanowski, J. S., Donatelli, M., Skidmore, A. K., & Vrieling, A. (2013). An auto-calibration procedure for empirical solar radiation models. *Environmental Modelling and Software*. <https://doi.org/10.1016/j.envsoft.2013.08.002>.
- Chen, R., Ersi, K., Yang, J., Lu, S., & Zhao, W. (2004). Validation of five global radiation models with measured daily data in China. *Energy Conversion and Management*. <https://doi.org/10.1016/j.enconman.2003.09.019>.
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*. <https://doi.org/10.1023/A:1022627411411>
- Da Silva, M. B. P., Francisco Escobedo, J., Juliana Rossi, T., dos Santos, C. M., & da Silva, S. H. M. G. (2017). 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*. <https://doi.org/10.1016/j.jastp.2017.04.001>.
- Dos Santos, C. M., Escobedo, J. F., Teramoto, É. T., & da Silva, S. H. M. G. (2016). Assessment of ANN and SVM models for estimating normal direct irradiation (H_b). *Energy Conversion and Management*. <https://doi.org/10.1016/j.enconman.2016.08.020>.
- Elizondo, D., Hoogenboom, G., & McClendon, R. W. (1994). Development of a neural network model to predict daily solar radiation. *Agricultural and Forest Meteorology*. [https://doi.org/10.1016/0168-1923\(94\)90103-1](https://doi.org/10.1016/0168-1923(94)90103-1).
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*.

- Hansen, J. W. (1999). Stochastic daily solar irradiance for biological modeling applications. *Agricultural and Forest Meteorology*. [https://doi.org/10.1016/S0168-1923\(99\)00003-9](https://doi.org/10.1016/S0168-1923(99)00003-9).
- Heinemann, A. B., van Oort, P. A. J., Fernandes, D. S., & Maia, A. de H. N. (2012). Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. *Bragantia*, 71(4), 572–582. <https://doi.org/10.1590/S0006-87052012000400016>.
- Hunt, L. A., Kuchar, L., & Swanton, C. J. (1998). Estimation of solar radiation for use in crop modelling. *Agricultural and Forest Meteorology*. [https://doi.org/10.1016/S0168-1923\(98\)00055-0](https://doi.org/10.1016/S0168-1923(98)00055-0).
- Jamieson, P. D., Porter, J. R., & Wilson, D. R. (1991). A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. *Field Crops Research*. [https://doi.org/10.1016/0378-4290\(91\)90040-3](https://doi.org/10.1016/0378-4290(91)90040-3).
- 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*. <https://doi.org/10.1016/j.rser.2015.08.021>.
- Landeras, G., López, J. J., Kisi, O., & Shiri, J. (2012). Comparison of Gene Expression Programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain). *Energy Conversion and Management*. <https://doi.org/10.1016/j.enconman.2012.03.025>.
- Lauret, P., Voyant, C., Soubdhan, T., David, M., & Poggi, P. (2015). A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Solar Energy*. <https://doi.org/10.1016/j.solener.2014.12.014>.
- Li, M. F., Tang, X. P., Wu, W., & Liu, H. Bin. (2013). General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Conversion and Management*. <https://doi.org/10.1016/j.enconman.2013.03.004>.
- Linares-Rodríguez, A., Ruiz-Arias, J.A., Pozo-Vázquez, & Tovar-Pescador, J. (2011). Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. *Energy*. <https://doi.org/10.1016/j.energy.2011.06.044>.
- Lorena, A. C., Jacintho, L. F. O., Siqueira, M. F., Giovanni, R. De, Lohmann, L. G., De Carvalho, A. C. P. L. F., & Yamamoto, M. (2011). Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2010.10.031>.
- Lyra, G. B., Zanetti, S. S., Santos, A. A. R., de Souza, J. L., Lyra, G. B., Oliveira-Júnior, J. F., & Lemes, M. A. M. (2016). Estimation of monthly global solar irradiation using the Hargreaves–Samani model and an artificial neural network for the state of Alagoas in northeastern Brazil. *Theoretical and Applied Climatology*. <https://doi.org/10.1007/s00704-015-1541-8>.

Marques Filho, E. P., Oliveira, A. P., Vita, W. A., Mesquita, F. L. L., Codato, G., Escobedo, J. F, França, J. R. A. (2016). Global, diffuse and direct solar radiation at the surface in the city of Rio de Janeiro: Observational characterization and empirical modeling Renewable Energy <https://doi.org/10.1016/j.renene.2016.01.040>.

Mohammadi, K., Shamshirband, S., Anisi, M. H., Amjad Alam, K., & Petković, D. (2015). Support vector regression based prediction of global solar radiation on a horizontal surface. Energy Conversion and Management. <https://doi.org/10.1016/j.enconman.2014.12.015>.

Mohammadi, K., Shamshirband, S., Tong, C. W., Arif, M., Petković, D., & Sudheer, C. (2015). A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation. Energy Conversion and Management. <https://doi.org/10.1016/j.enconman.2014.12.050>.

Platt, J. (1999). Fast Training of Support Vector Machines using Sequential Minimal Optimization. In *Advances in Kernel Methods - Support Vector Learning*.

Podestá, G. P., Núñez, L., Villanueva, C. A., & Skansi, M. A. (2004). Estimating daily solar radiation in the Argentine Pampas. *Agricultural and Forest Meteorology*. <https://doi.org/10.1016/j.agrformet.2003.11.002>.

Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews*. <https://doi.org/10.1016/j.rser.2014.07.108>.

Rehman S, Mohandes M. Splitting global solar radiation into diffuse and direct normal fractions using artificial neural networks. *Energy Sources, Part A* 2012;34:1326–36.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., & Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks*. <https://doi.org/10.1109/72.870050>.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.

Soares, P. L. B., & Da Silva, J. P. (2011). Aplicação de Redes Neurais Artificiais em Conjunto com o Método Vetorial da Propagação de Feixes na Análise de um Acoplador Direcional Baseado em Fibra Ótica. *Revista Brasileira de Computação Aplicada*. <https://doi.org/10.5335/rbca.2011.1803>.

Stone, R.J.. Improved statistical procedure for the evaluation of solar radiation estimation models. *Solar Energy*, vol. 51, no 4, p. 289 – 291, 1993.

Weiss, A., & Hays, C. J. (2004). Simulation of daily solar irradiance. *Agricultural and Forest Meteorology*. <https://doi.org/10.1016/j.agrformet.2003.12.002>.

Yadav, A. K., & Chandel, S. S. (2014). Solar radiation prediction using Artificial Neural Network techniques: A review. *Renewable and Sustainable Energy Reviews*. <https://doi.org/10.1016/j.rser.2013.08.05>.