



**UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL**

CHARLES CAMPOE MARTIM

**MODELOS EMPÍRICOS E MACHINE LEARNING NA ESTIMATIVA DA
RADIAÇÃO GLOBAL DIÁRIA NA AMAZÔNIA**

**CUIABÁ - MT
2024**

CHARLES CAMPOE MARTIM

**MODELOS EMPÍRICOS E MACHINE LEARNING NA ESTIMATIVA DA
RADIÇÃO GLOBAL DIÁRIA NA AMAZÔNIA**

Tese apresentada ao Programa de Pós-Graduação em Física Ambiental da Instituto de Física da Universidade Federal de Mato Grosso como requisito parcial para a obtenção do título de Doutor em Física Ambiental.

Orientador: Prof. Dr. Adilson Pacheco de Souza

**CUIABÁ - MT
2024**

Dados Internacionais de Catalogação na Fonte.

M378m Martim, Charles Campoe.

Modelos empíricos e machine learning na estimativa da radiação global diária na Amazônia [recurso eletrônico] / Charles Campoe Martim. -- Dados eletrônicos (1 arquivo : 110 f., il. color., pdf). -- 2024.

Orientador: Adilson Pacheco de Souza.

Tese (doutorado) - Universidade Federal de Mato Grosso, Instituto de Física, Programa de Pós-Graduação em Física Ambiental, Cuiabá, 2024.

Modo de acesso: World Wide Web: <https://ri.ufmt.br>.

Inclui bibliografia.

1. SVM. 2. ANN. 3. Bioma amazônico. 4. Predição. 5. Modelos simplificados. I. Souza, Adilson Pacheco de, *orientador*. II. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Permitida a reprodução parcial ou total, desde que citada a fonte.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE MATO GROSSO
PRÓ-REITORIA DE ENSINO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

FOLHA DE APROVAÇÃO

TÍTULO: MODELOS EMPÍRICOS E MACHINE LEARNING NA ESTIMATIVA DA RADIAÇÃO GLOBAL DIÁRIA NA AMAZÔNIA

AUTOR: DOUTORANDO CHARLES CAMPOE MARTIM

Tese defendida e aprovada em **25 de OUTUBRO de 2024**.

COMPOSIÇÃO DA BANCA EXAMINADORA

1. PROF. DR. ADILSON PACHECO DE SOUZA (Presidente Banca / **ORIENTADOR**)

INSTITUIÇÃO: UNIVERSIDADE FEDERAL DE MATO GROSSO/SINOP

2. PROFA. DRA. DANIELA ROBERTA BORELLA (Membro Externo)

INSTITUIÇÃO: FACULDADE DE TECNOLOGIA DE SINOP/FASTECH

3. PROF. DR. FREDERICO TERRA DE ALMEIDA (Membro Externo)

INSTITUIÇÃO: INSTITUTO DE CIÊNCIAS AGRÁRIAS E AMBIENTAIS/UFMT

4. PROF. DR. JOÃO GABRIEL RIBEIRO DAMIAN (Membro Externo)

INSTITUIÇÃO: FACULDADE DE CIÊNCIAS EXATAS E TECNOLÓGICAS/UNEMAT/SINOP

5. PROF. DR. ÉRICO TADAO TERAMOTO (Membro Externo)

INSTITUIÇÃO: FACULDADE DE CIÊNCIAS AGRÁRIAS DO VALE DO RIBEIRA/
UNESP/ Registro

CUIABÁ, 25/10/2024.



Documento assinado eletronicamente por **MARCELO SACARDI BIUDES, Coordenador(a) de Programas de Pós-Graduação em Física Ambiental - IF/UFMT**, em 29/10/2024, às 09:32, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **João Gabriel Ribeiro Damian, Usuário Externo**, em 29/10/2024, às 09:45, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Érico Tadao Teramoto, Usuário Externo**, em 29/10/2024, às 10:07, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **ADILSON PACHECO DE SOUZA, Docente da Universidade Federal de Mato Grosso**, em 29/10/2024, às 10:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Daniela Roberta Borella, Usuário Externo**, em 29/10/2024, às 10:16, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **FREDERICO TERRA DE ALMEIDA, Docente da Universidade Federal de Mato Grosso**, em 29/10/2024, às 10:28, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufmt.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7234948** e o código CRC **FB28440F**.

DEDICATÓRIA

Aos meus pais Santo Pigosso Martim e Rosilene Campoe Martim e meus irmãos Alini Campoe Martim e Marco Aurélio Campoe Martim pelo apoio incondicional ao longo de toda a minha trajetória no Doutorado.

AGRADECIMENTOS

Ao Professor Dr. Adilson Pacheco de Souza pela orientação durante toda a execução desta tese, com auxílio, apontamentos e direcionamento que foram de extrema importância para a finalização deste trabalho.

A Universidade Federal de Mato Grosso por todo o suporte fornecido.

Ao Grupo de Pesquisa Tecnologia em Recursos Hídricos no Centro-Oeste (TREHCO), em especial ao Rhavel Salviano, Daniela Castagna, Luzinete, Célio e Jones pela amizade e contribuição na elaboração desta tese.

A banca examinadora, composta pelos professores doutores Daniela, Frederico, João Gabriel e Érico Tadao pelo tempo despendido na avaliação e melhorias deste trabalho.

Ao Programa de Pós-Graduação em Física Ambiental, pela oportunidade de estar cursando o Doutorado, em especial a todos os professores do programa, pelos ensinamentos ao longo desse período.

A Coordenação de Aperfeiçoamento de Nível Superior (CAPES) pelo apoio financeiro através da concessão da bolsa de Doutorado, Código de Financiamento – 001.

Aos meus familiares pelo apoio emocional, que foi primordial nesse período.

E a todas as pessoas que contribuíram de forma direta ou indireta na execução deste trabalho.

RESUMO

MARTIM, C. C. MODELOS EMPÍRICOS E MACHINE LEARNING NA ESTIMATIVA DA RADIAÇÃO GLOBAL DIÁRIA NA AMAZÔNIA, 127f. Tese (Doutorado em Física Ambiental) – Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, 2024.

A radiação global (Hg) influencia diretamente vários processos químicos, físicos e biológicos entre eles está a evapotranspiração, fotossíntese, geração de energia fotovoltaica entre outros, porém sua mensuração está restrita a certas regiões, sendo necessário a série histórica para estudos científicos, projetos agrícolas, ambientais e industriais. Algumas variáveis meteorológicas como a temperatura do ar ou insolação apresentam uma boa correlação com a Hg. Neste trabalho, foram avaliados 87 modelos empíricos para a estimativa da radiação global com insolação, temperatura do ar, umidade relativa do ar, variáveis astronômicas e combinações híbridas com mais de uma variável, além disso, foi avaliado as técnicas de machine learning do tipo MLP e SVM com 40 diferentes combinações de entrada, para 20 cidades distribuídas no Bioma Amazônico brasileiro. Com os modelos empíricos, empregando a insolação, o melhor desempenho foi obtido com o modelo potencial, com S, So e Ho como variável de entrada, já para a temperatura do ar, o melhor desempenho foi obtido com o modelo que possui como variável de entrada o ΔT , Tmed e Ho, já o modelo híbrido com ΔT , S, So e Ho como variável de entrada, e obteve o melhor desempenho, já os modelos com variáveis astronômicas foram não significativos (NS) para todas as estações meteorológicas avaliadas e os modelos baseados na umidade relativa do ar apresentaram baixo desempenho. Na avaliação por meio do MLP e SVM, quanto maior o número de variáveis de entrada, melhor o desempenho, principalmente com a inserção da insolação e temperatura do ar, que reduz o espalhamento em dias com alta e baixa transmissividade atmosférica, porém quando tem a disponibilidade de todas as variáveis é recomendado o emprego da combinação RHmax, RHmed, RHmin, Tmax, Tmed, Tmin, S, So, Ho. No geral, não houve diferença expressiva no desempenho dos modelos empíricos e dos MLP e SVM com dados meteorológicos locais para a estimativa da radiação global no bioma Amazônico brasileiro.

Palavras-chave: SVM; ANN; bioma Amazônico; Predição, Modelos simplificados.

ABSTRACT

MARTIM, C. C. EMPIRICAL MODELS AND MACHINE LEARNING IN ESTIMATING DAILY GLOBAL RADIATION IN THE AMAZON, 127f. Thesis (Doctor in Environmental Physics) – Physics of Institute, Federal University of Mato Grosso, Cuiabá, 2024.

Global radiation (Hg) directly influences various chemical, physical and biological processes, including evapotranspiration, photosynthesis, photovoltaic energy generation and others. However, its measurement is restricted to certain regions and historical series are necessary for scientific studies, agricultural, environmental and industrial projects. Some meteorological variables such as air temperature or insolation correlate well with Hg. This study evaluated 87 empirical models for estimating global radiation using insolation, air temperature, relative humidity, astronomical variables and hybrid combinations with more than one variable, and also evaluated MLP and SVM machine learning algorithms with 40 different input combinations, for 20 cities distributed in the Brazilian Amazon Biome. With the empirical models, using insolation, the best performance was obtained with the potential model, with S, So and Ho as input variables, while for air temperature, the best performance was obtained with the model with ΔT , Tmed and Ho as input variables, The hybrid model with ΔT , S, So and Ho as input variables had the best performance, while the models with astronomical variables were non-significant (NS) for all the weather stations evaluated and the models based on relative humidity showed poor performance. In the evaluation using the MLP and SVM algorithms, the greater the number of input variables, the better the performance, especially with the inclusion of insolation and air temperature, which reduces scattering on days with high and low atmospheric transmissivity, but when all the variables are available it is recommended to use the combination RHmax, RHmed, RHmin, Tmax, Tmed, Tmin, S, So, Ho. Overall, there was no significant difference in the performance of the empirical models and the MLP and SVM algorithms with local meteorological data for estimating global radiation in the Brazilian Amazon biome.

Keywords: SVM; ANN; Amazon biome; Prediction, Simplified models.

LISTA DE ILUSTRAÇÕES

Figura 2.1. Precipitação pluvial anual para o Brasil (Adaptado de Alvares et al. 2013).	6
Figura 2.2 Classificação Climática de Köppen para o Brasil (Adaptado de Alvares et al. 2013).....	7
Figura 2.3. Inteligência artificial, aprendizado de máquina, aprendizado supervisionado, regressão, classificação (Adaptado de Agbulut et al. 2021).	16
Figura 2.4. Estrutura básica de uma rede neural artificial (RNA) (Adaptado de Santos et al. 2021).....	18
Figura 2.5. SVM de a) classificação (classe 1 e 2) e b) regressão (Adaptado de Bellido-Jiménez et al. 2021).	19
Figura 3.1. Localização das 20 estações meteorológicas avaliadas neste estudo, no bioma Amazônia, Brasil. (A ordem numérica das estações está apresentada na Tabela 1).	37
Figura 3.2. Correlações, distribuições de frequência e interpolação entre as variáveis meteorológicas (Hg, Tmax, Tmin, Tmed, RHmed, S) no bioma Amazônico brasileiro.....	50
Figura 3.3. Boxplots da radiação global medida (Hg) e da radiação extraterrestre calculada (Ho) para 20 estações meteorológicas na Amazônia brasileira, em agrupamentos de dados mensal.	50
Figura 3.4. Variação da radiação global diária (Hg) e extraterrestre (Ho) para as estações meteorológicas de Boa Vista, Manaus e Sinop, na Amazônia brasileira, ao longo de diferentes anos de medidas.....	51
Figura 3.5. Boxplots dos indicativos de desempenho estatístico (R^2 , MBE, RMSE e d) para 87 modelos simplificados de estimativas de Hg, considerando os valores agrupados de 20 estações meteorológicas do bioma Amazônico brasileiro. Em que: coeficiente de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE) e índice de concordância de Willmott (d).	53
Figura 3.6. Indicativos de desempenho estatístico dos melhores modelos simplificados de estimativas de cada grupo (modelos 10, 45, 46 e 62) para 20 estações meteorológicas do bioma Amazônico brasileiro. Em que coeficiente de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE) e índice de concordância de Willmott (d).....	55
Figura 3.7. Dispersão entre a radiação global medida e estimada pelos modelos 10, 45 e 62 para as estações meteorológicas de Boa Vista (RR), Manaus (AM) e Sinop (MT), em diferentes classes de transmissividade atmosférica. (A linha cinza, representa a relação 1:1 ou $y = x$).	57
Figura 3.8. Frequência do erro absoluto nas estações meteorológicas de Boa Vista, Manaus e Sinop, para os modelos 10, 45 e 62 em diferentes transmissividades atmosférica.	58
Figura 3.9. Teste-t dos coeficientes dos modelos, com 0,05 de significância	72

Figura 3.10. Frequência acumulada do erro absoluto dos modelos 10, 45 e 62 nas 20 estações meteorológicas localizadas no Bioma Amazônico brasileiro.....	73
Figura 4.1. Mapa de localização das 20 estações meteorológicas automáticas e convencionais distribuídas pelo bioma Amazônia.....	79
Figura 4.2. Estrutura organizacional da rede neural artificial desenvolvida para MLP.....	84
Figura 4.3. Máquina de vetores de suporte (SVM).....	85
Figura 4.4. Correlação de Pearson (r) entre as variáveis meteorológicas obtidas pelas 20 EMAs e EMCs localizadas no bioma Amazônico brasileiro.....	90
Figura 4.5. Boxplot dos coeficientes de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE), erro absoluto médio (MAE) e coeficiente de Willmott (d), para o MLP em 40 combinações de variáveis de entrada, em 20 estações meteorológicas do bioma Amazônico brasileiro. (As diferentes cores representam os agrupamentos das combinações).....	93
Figura 4.6. Boxplot dos coeficientes de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE), erro absoluto médio (MAE) e coeficiente de Willmott (d), para o SVM em 40 combinações de variáveis de entrada, em 20 estações meteorológicas do bioma Amazônico brasileiro. (As diferentes cores representam os agrupamentos das combinações).....	93
Figura 4.7. Dispersão entre valores medidos e estimados de Hg pelo MLP em diferentes condições de cobertura de céu, nas estações meteorológicas de Boa Vista, Manaus e Sinop, na Amazônia brasileira.....	96
Figura 4.8. Dispersão entre valores medidos e estimados de Hg pelo SVM em diferentes condições de cobertura de céu, nas estações meteorológicas de Boa Vista, Manaus e Sinop, na Amazônia brasileira.....	96
Figura 4.9. Valores médios do coeficiente de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro médio (RMSE), erro médio absoluto (MAE) e índice de concordância de Willmott (d), para o MLP (branco) e SVM (cinza), em diferentes combinações de variáveis de entrada (1, 7 e 40).....	97
Figura 4.10. Dispersão entre radiação global medida e radiação global estimada por MLP, SVM e modelos empíricos simplificados, para as estações meteorológicas de Boa Vista-RR, Manaus-AM e Sinop-MT, considerando diferentes variáveis de entrada dos modelos.....	99
Figura 4.11. Frequência do erro absoluto na estimativa da radiação global empregando modelo empírico e técnicas de ML com a insolação como variável de entrada (A), temperatura do ar (B) e combinação híbrida (C) como variável entrada para as estações meteorológicas de Boa Vista, Manaus e Sinop.....	100

LISTA DE TABELAS

Tabela 2.1. Parâmetros C , γ e ϵ utilizados no SVM para estimativa da radiação global.	21
Tabela 2.2. Parâmetros dos modelos de rede neural artificial (ANN).	22
Tabela 2.3. Combinações de entrada utilizada em vários trabalhos com técnicas de ML.	22
Tabela 2.4. Combinações de entrada para ANN e SVM (Adaptado de Santos et al. 2021).	23
Tabela 3.1. Estações meteorológicas na Amazônia brasileira e suas respectivas localizações, classificação climática e período de dados.	37
Tabela 3.2. Modelos empíricos de estimativas indiretas de radiação global baseados na insolação.	40
Tabela 3.3. Modelos empíricos de estimativas diárias da radiação global baseados na temperatura do ar.	41
Tabela 3.4. Modelos empíricos de estimativas diária da radiação global baseados na umidade relativa do ar (RHmed) e com variáveis astronômicas.	42
Tabela 3.5. Modelos empíricos híbridos de estimativas diárias da radiação global com variáveis meteorológicas e astronômicas.	43
Tabela 3.6. Médias anuais e desvio padrão das variáveis meteorológicas (Hg, S, Tmax, Tmed, Tmin, RHmed e Prec.) e astronômica (Ho) para as 20 estações meteorológicas estudadas na Amazônia brasileira (para precipitação pluvial considera-se os totais anuais médios).	47
Tabela 3.7. Coeficientes calibrados para os modelos 10, 45 e 62 de estimativas da radiação global diária, para cada estação meteorológica avaliada no Bioma Amazônico.	58
Tabela 4.1. Estações meteorológicas automáticas inseridas na Amazônia brasileira.	79
Tabela 4.2. Diferentes combinações das variáveis de entrada para estimativas de Hg pelo MLP e SVM, no Bioma Amazônico brasileiro.	86
Tabela 4.3. Médias diárias anuais das variáveis meteorológicas e empíricas para as 20 estações meteorológicas avaliadas no Bioma Amazônico.	91

LISTA DE ABREVIATURAS E SIGLAS

ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Rede neural artificial
BRF	Base Radial de Kernel
C	Custo
CE	Camada de Entrada
CO	Camada Oculta
CS	Camada de Saída
DJ	Dia Juliano
DL	Deep Learning
d	Índice de concordância de Willmott
EMCs	Estações meteorológicas convencionais
EMAs	Estações meteorológicas automáticas
ELM	Extreme Learning Machine
FA	Função de Ativação
GRNN	Generalized Regression Neural Network
GP	Genetic Programming
GPML	Gaussian Processes of Machine Learning
GBDT	Gradiente Boosting Decision Tree
Hg	Radiação global ($\text{MJ m}^{-2} \text{ d}^{-1}$)
Ho	Radiação Extraterrestre ($\text{MJ m}^{-2} \text{ d}^{-1}$)
Hc	Radiação global em um dia completamente claro ($\text{MJ m}^{-2} \text{ d}^{-1}$)
IT	Número de Interações
INMET	Instituto Nacional de Meteorologia
K-NN	Kernel Nearest Neighbor
LSSVM	Least-Square Support Vector Machine
MARS	Multivariate Adaptive Regression Spline
ML	Aprendizado de Máquina
MO	Momentum
M5Tree	M5 model Tree
MLP	Perceptron Multi-Camadas
MBE	Erro absoluto médio ($\text{MJ m}^{-2} \text{ d}^{-1}$)
Prec.	Precipitação (mm)
R	Coefficiente de correlação
R ²	Coefficiente de determinação
RMSE	Raiz do erro quadrático médio ($\text{MJ m}^{-2} \text{ d}^{-1}$)
RBFNN	Radial Basis Function Neural Network
RF	Random Forest
RHmax	Umidade relativa máxima

RHmed	Umidade relativa média
RHmin	Umidade relativa mínima
So	Fotoperíodo (horas)
S	Insolação (horas)
SVM	Máquina de Vetores de Suporte
SVR	Support Vector Regression
TA	Taxa de Aprendizagem
Tmax	Temperatura máxima
Tmed	Temperatura média
Tmin	Temperatura mínima
Vp	Pressão de vapor de água
VCE	Variáveis da Camada de Entrada
Vv	Velocidade do Vento (m s^{-1})
XGBoost	Extreme Gradient Boost

LISTA DE SÍMBOLOS

ϵ	Épsilon
γ	Gama
β	Viés de cada camada oculta

SUMÁRIO

1.0 INTRODUÇÃO GERAL	1
1.1 Problemática	1
1.2 Justificativa	2
1.3 OBJETIVOS	4
1.3.1 Objetivo Geral	4
1.3.2 Objetivos Específicos	4
2.0 FUNDAMENTAÇÃO TEÓRICA	5
2.1 Bioma Amazônia	5
2.2 MODELOS EMPÍRICOS	7
2.2.1 Modelos baseados na insolação	7
2.2.2 Modelos baseados na temperatura do ar	11
2.2.3 Modelos híbridos	13
2.2.4 Modelos baseados em variáveis astronômicas	14
2.3 INTELIGÊNCIA ARTIFICIAL (IA) E MACHINE LEARNING (ML)	15
2.3.1 Rede Neural Artificial (ANN)	16
2.3.2 Suporte de vetor e máquina (SVM)	18
2.3.3 Estimativa da radiação global com MLP e SVM	20
2.4 REFERÊNCIAS BIBLIOGRÁFICAS	26
3.0 CAPÍTULO 1 - ESTIMATIVAS DIÁRIAS DA RADIAÇÃO GLOBAL NA AMAZÔNIA BRASILEIRA POR MODELOS SIMPLIFICADOS	30
3.1 INTRODUÇÃO	30
3.2 MATERIAL E MÉTODOS	35
3.2.1 Área de estudo	35
3.2.2 Modelos simplificados da estimativa de HG	37
3.2.3 Modelos baseados na insolação – GRUPO I	39
3.2.4 Modelos baseados na temperatura do ar – GRUPO II	40
3.2.5 Modelos baseados na umidade relativa do ar ou em variáveis astronômicas – GRUPO III e IV	42
3.2.6 Modelos empíricos híbridos – GRUPO V	42
3.2.7 Desempenho estatístico dos modelos	43
3.3 RESULTADOS	45
3.4 DISCUSSÃO	59
3.5 CONCLUSÃO	64
3.6 REFERÊNCIAS BIBLIOGRÁFICAS	66
3.9 ANEXO	72

4.0 CAPÍTULO 2 - MACHINE LEARNING NA ESTIMATIVA DA RADIAÇÃO GLOBAL DIÁRIA NA AMAZÔNIA BRASILEIRA.....	74
4.1 INTRODUÇÃO.....	74
4.2 MATERIAL E MÉTODOS.....	78
4.2.1 Área de estudo	78
4.2.2 Análise dos dados.....	80
4.2.3 Inteligência artificial (IA) e Machine Learning (ML).....	81
4.2.3.1 Rede Neural Artificial (ANN) do tipo Multi Layer Perceptron (MLP)	82
4.2.3.2 Máquina de Vetores de Suporte (SVM)	84
4.2.4 Estrutura dos modelos de ML	86
4.2.5 Modelos empíricos	87
4.2.6 INDICATIVOS ESTATÍSTICOS.....	88
4.3 RESULTADOS.....	89
4.4 DISCUSSÃO.....	100
4.5 CONCLUSÃO	105
4.6 REFERÊNCIAS BIBLIOGRÁFICAS	106
5.0 CONSIDERAÇÃO FINAL E GERAL	110

1.0 INTRODUÇÃO GERAL

1.1 Problemática

A falta de dados meteorológicos, é um desafio significativo em todos os países, especialmente em regiões remotas ou países em desenvolvimento, principalmente por falta de infraestrutura, alto custo de manutenção, necessidade de pessoas capacitadas, atualização tecnológica das estações meteorológicas entre outros fatores. Dados meteorológicos confiáveis são cruciais em várias atividades como, modelagem de condições climáticas, planejamento e ações adotadas em desastres naturais, planejamento de recursos hídricos, planejamento agrícola e estudos relacionados a influência das mudanças climáticas locais.

Embora no Brasil, parte significativa dos dados meteorológicos disponíveis para estudos são das estações meteorológicas automáticas (EMAs) e convencionais (EMCs) são geridas pelo Instituto Nacional de Meteorologia (INMET), e estão distribuídas espacialmente em todas as unidades da federação, porém são insuficientes para o estudos meteorológicos regionais em muitos locais do país, principalmente em função da distribuição espacial desigual, pois a maior parte dessas estações meteorológicas estão concentradas nas regiões Sul, Sudeste, Nordeste e Centro-Oeste, enquanto na região Norte que abrange boa parte do bioma Amazônico, ocorre a menor cobertura espacial das EMAs e EMCs em todo o país.

O bioma Amazônico brasileiro é uma região estratégica em função de sua grande dimensão, pois ocupa aproximadamente metade do território brasileiro, sendo caracterizado pelo clima Tropical, com alto volume de precipitação pluvial e baixa variação de fotoperíodo, o que permite a manutenção de uma grande diversidade de flora e de fauna. A região é compreendida pela bacia Amazônica, onde o principal rio é o Amazonas, que possui inúmeros afluentes ao longo de sua extensão.

É uma região com contribuição significativa na atividade econômica do Brasil, com a pecuária sendo a principal, porém recentemente a agricultura vem se expandindo, tornando essa região a nova fronteira agrícola brasileira, com a implantação de extensas áreas de lavoura, principalmente soja e milho. A região também apresenta um grande potencial para a produção de energia fotovoltaica, sendo necessários projetos para avaliar a viabilidade de produção em cada local, através das

medições *in loco* da radiação global (Hg). Sendo que o conhecimento da radiação global incidente é um fator crucial em todos os setores, principalmente na produção agropecuária, pois impacta no crescimento das plantas, na eficiência do uso de recursos como a água em sistemas de irrigação e o bem-estar animal.

Por ser uma região próxima a Linha do Equador (latitude de 5 a -15°), a radiação extraterrestre (H_o) nesta região é praticamente constante ao longo do ano, porém ao atravessar a atmosfera, a radiação solar acaba interagindo com diferentes constituintes atmosféricos, e a radiação global incidente pode apresentar mudança significativa, tanto espacial como temporal.

No monitoramento das variáveis meteorológicas, o bioma Amazônico possui apenas 72 EMAs, que ainda estão concentradas em poucas cidades, em razão de diversos fatores, como a difícil logística da região para a instalação, coleta de dados e a manutenção que acaba sendo mais onerosa. Além disso, parte da série histórica das estações meteorológicas já consolidadas, apresentam dados falhos ou inconsistentes significativos, principalmente para a Hg, que segundo Martim & Souza (2021) está entre 11 e 59% e que para Souza et al. (2017) fica entre 14,20 e 59,28% de falhas da série histórica, que ocorre principalmente em razão da falta de manutenção periódica, principalmente o sensor piranômetro que necessita de manutenção especializada periodicamente.

1.2 Justificativa

Pesquisadores de diferentes países, vem se dedicando a estudos que visam a avaliação das diferentes metodologias empregadas na estimativa da Hg, sendo assim, todos os anos vários artigos são publicados envolvendo a temática, como descrito na revisão de Prieto & García (2022) onde os autores encontraram 165 modelos empíricos, e na revisão de Zhou et al (2021) os autores encontraram 232 artigos que empregaram diferentes técnicas de Machine Learning (ML) na estimativa da Hg, com diversas variáveis de entrada.

Apesar dos dados da radiação global ser de difícil acesso, seja por falhas na série histórica ou falta de medição na região de interesse, existem variáveis meteorológicas que possuem boa correlação com a radiação global, e são de fácil

medição, como é o caso da temperatura do ar e ainda em algumas EMCs é realizado a medição da insolação.

Quando se trata de modelagem da radiação global, algumas metodologias são mais utilizadas, como é o caso do modelo de Angström-Prescott que correlaciona linearmente a radiação global com a insolação, o modelo de Bristow & Campbell e Hargreaves & Samani que correlacionam a radiação global com a temperatura do ar, e atualmente vem sendo empregado também com bom desempenho as técnicas de Machine Learning, entre eles o mais utilizado é a rede neural artificial (ANN) e o suporte de vetor e máquina (SVM) para o preenchimento de falhas na base de dados. Os modelos empíricos são baseados em relações matemáticas ou estatísticas, desenvolvidas a partir de dados observacionais ou experimentais, ao invés de serem baseados em princípios teóricos ou leis fundamentais, já as técnicas de ML conseguem através dos dados de entrada, identificar padrões e fazer previsões futuras, mesmo em problemas não lineares. Na estimativa da radiação global, pode-se empregar diversas variáveis de entrada como a insolação, temperatura do ar, umidade relativa do ar, variáveis astronômicas e combinações híbridas, porém independente da metodologia adotada, é necessário calibrar os modelos empíricos e treinar as técnicas de Machine Learning e posteriormente fazer a validação a fim de avaliar o desempenho nas condições meteorológicas locais.

No bioma Amazônico, há poucos artigos que avaliaram o desempenho destas metodologias de uma forma mais abrangente, sendo o trabalho mais completo publicado por Martim & Souza (2021) que só avaliaram o modelo de Angstrom-Prescott para 20 locais, enquanto Souza et al. (2017) empregaram 15 modelos baseados na temperatura do ar e precipitação pluvial para 14 cidades no estado de Mato Grosso, e no estudo de Ramos et al. (2018) que avaliaram os modelos de Bristow & Campbell e Hargreaves & Samani, e que apesar de abranger todo o Brasil com 32 locais, apenas 4 estão inseridas no bioma Amazônia. Já relacionado as técnicas de Machine Learning para o Bioma Amazônico, o único trabalho desenvolvido foi publicado por Marques et al. (2024) que avaliaram as estimativas da radiação global em 12 locais, inseridas no estado do Amazonas, onde os autores avaliaram apenas uma combinação de variáveis de entrada, sendo recomendado pelos autores, que em trabalhos futuros seja avaliado diferentes combinações de variáveis meteorológicas de entrada.

A necessidade de estudos mais abrangentes para a região do bioma Amazônico, envolvendo os diferentes métodos de estimativas da Hg, tanto os modelos empíricos como as técnicas de ML, principalmente na avaliação do desempenho e na seleção de metodologias mais viáveis, porém sempre considerando as variáveis meteorológicas mais facilmente disponíveis.

O que se observa, é que essas metodologias de estimativa da radiação global, já estão consolidadas em vários países, porém principalmente na região que compreende o bioma Amazônico, existem poucos estudos, e que foram desenvolvidos para locais específicos, sendo que existe a necessidade de estudos mais abrangentes sobre este tema, para uma compreensão mais ampla do potencial dessas diferentes metodologias.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Estimar a radiação global (Hg) por meio de modelos empíricos e as técnicas de Machine Learning e avaliar as melhores metodologias para aplicações ambientais, energéticas e agrícolas no bioma Amazônico.

1.3.2 Objetivos Específicos

Revisão de literatura e avaliação do desempenho dos modelos empíricos com diferentes variáveis de entrada, em estações meteorológicas inseridas no bioma Amazônico brasileiro.

Avaliar a capacidade preditiva das técnicas de machine learning (ML) do tipo rede neural artificial (ANN) e suporte de vetor e máquina (SVM) com diferentes combinações de variáveis de entrada, e também comparar com o desempenho dos melhores modelos empíricos selecionados para a estimativa da radiação global no bioma Amazônico brasileiro.

2.0 FUNDAMENTAÇÃO TEÓRICA

2.1 Bioma Amazônia

O bioma Amazônico do Brasil, abrange uma extensa área de aproximadamente 4.196.943 km² que corresponde a 49% do território brasileiro, e está localizado entre as latitudes 5 a -15°. Por ser uma região próxima a Linha do Equador, a radiação extraterrestre (H_o) e o fotoperíodo (S_o) em média ao longo do ano é de 36 MJ m⁻² d⁻¹ e 12 horas, porém a radiação global (H_g) e a insolação (S) sofrem grandes mudanças ao longo do ano, como demonstrado por Martim et al. (2020) que interfere diretamente na transmissividade atmosférica local, pois nesta região, como observado na Figura 2.1, algumas cidades, como por exemplo a São Gabriel da Cachoeira no estado do Amazonas a precipitação ultrapassa os 3.100 mm anuais, em Sinop no estado de Mato Grosso ficou entre 1.900 a 2.200 segundo Alvares et al. (2013), influenciando diretamente a cobertura do céu com a presença de nuvens e mudanças da composição atmosférica, resultando em grande variabilidade espaço-temporal da radiação global (SOUZA et al. 2017).

Embora a região apresente alta precipitação pluvial anual, que reduz a transmissividade atmosférica em média de 50%, ocorre que na região, a radiação global incidente durante o ano, apresenta altos valores, que variam de 16,55 a 20,33 MJ m⁻² d⁻¹ na média mensal, segundo Martim & Souza (2021), em função da sua posição geográfica, próximo a Linha do Equador.

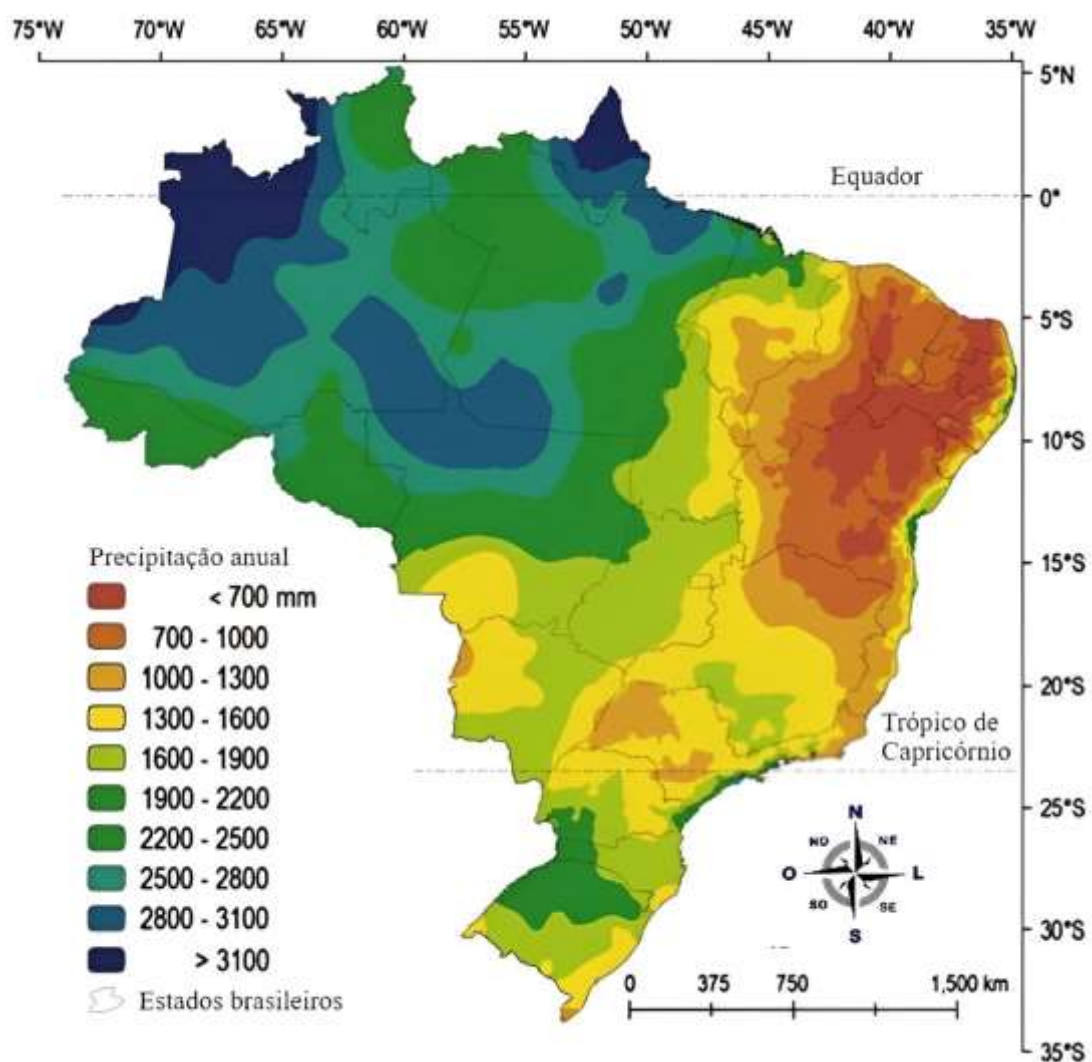


Figura 2.1. Precipitação pluvial anual para o Brasil (Adaptado de Alvares et al. 2013).

Segundo Alvares et al. (2013) 81% do território brasileiro se enquadra na Zona Climática Tropical, representado pela letra A, que é subdividido em estação seca (Af), monção (Am), com verão seco (As) e inverno seco (Aw), sendo que boa parte desta área contempla o bioma Amazônico, como mostrado na Figura 2.2, e esta caracterização deve-se a diversos fatores comuns nesta grande extensão territorial, como variações nas altitudes, precipitações pluviais e temperaturas do ar que condicionam essa zona climática.

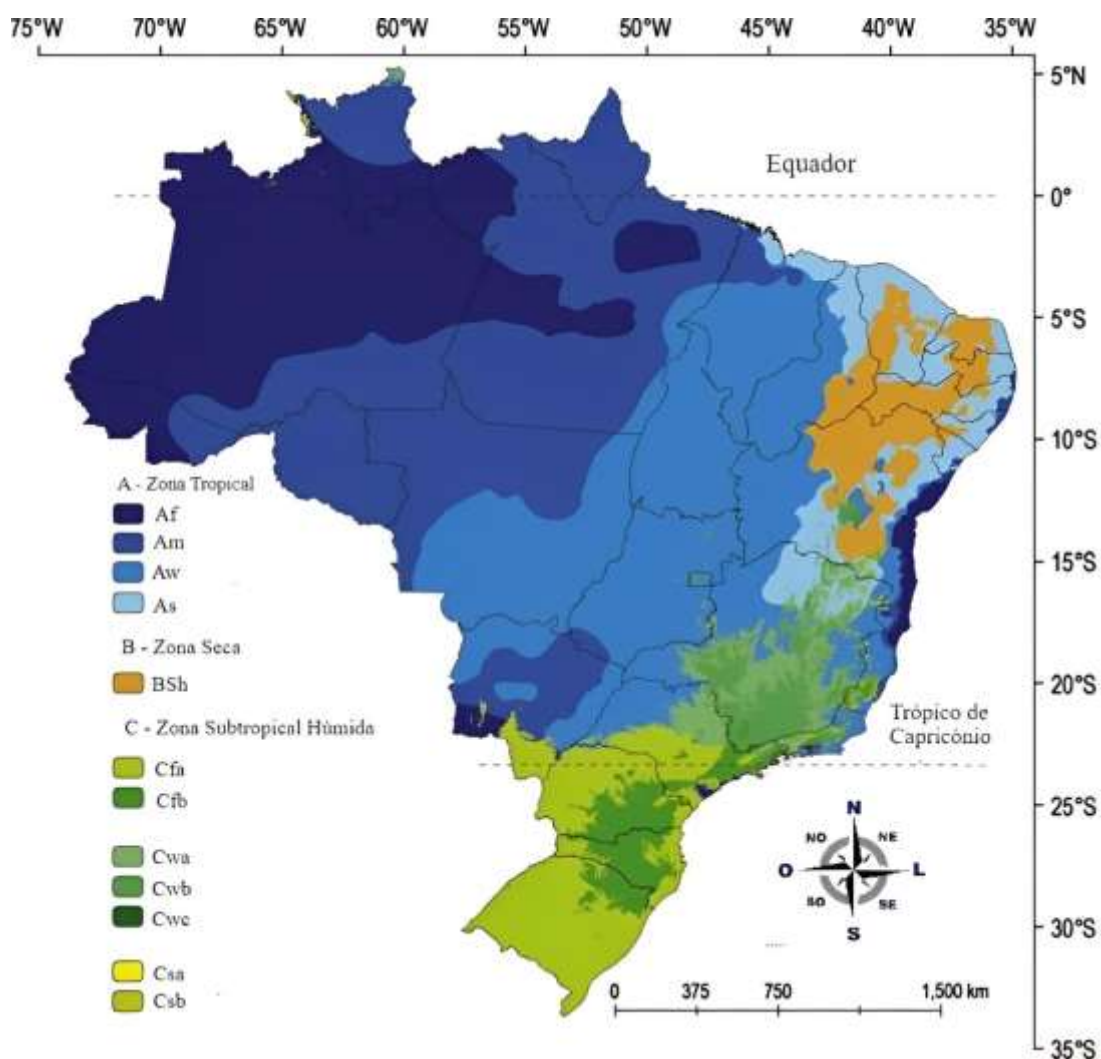


Figura 2.2 Classificação Climática de Köppen para o Brasil (Adaptado de Alvares et al. 2013).

2.2 MODELOS EMPÍRICOS

2.2.1 Modelos baseados na insolação

A insolação (S) é uma medida do tempo de incidência da radiação direta em horas sobre um determinado local, sendo o valor diário da S menor que o fotoperíodo (S_0) que é um valor teórico ($S < S_0$), pois não sofre a influência dos componentes atmosféricos, já quando o dia está completamente nublado a $S = 0$, porém em dias normais a $S > 0$. A insolação (S) é medida no Heliógrafo de Campbell-Stokes, que possui uma esfera de vidro, onde os raios solares são direcionados ao heliograma, que é uma fita de papel com gramatura e cor padronizada, impressa com subdivisão de 0,5

horas, onde o registro ocorre quando a fita é perfurada com a incidência da radiação direta ao longo do dia.

Utilizando a radiação global (H_g), radiação para um dia completamente claro (H_c), insolação (S) e o fotoperíodo (S_o) foi proposto por Angström (1924) para a cidade de Estocolmo, para um dia completamente claro o modelo empírico recomendado foi $H_g/H_c = 0,25 + 0,75 (S / S_o)$ que relaciona linearmente a transmissividade atmosférica (H_g / H_c) e a razão de insolação (S / S_o), sendo observado por Prescott (1940) a limitação do emprego do modelo proposto na maioria dos dias, sendo assim, foi proposto a substituição do H_c pela radiação extraterrestre (H_o), e empregando dados do Monte Stramlo, na cidade de Camberra localizada na Austrália, foi proposto que a radiação global poderia ser estimada pela equação $H_g / H_o = 0,25 + 0,54 (S / S_o)$. Sendo que a transmissividade atmosférica está diretamente relacionada as condições atmosféricas, após o ano de 1940, vários trabalhos foram publicados ajustando os coeficientes do modelo conhecido como Angstrom-Prescott em homenagem os pesquisadores.

Por mais que os modelos forneçam os valores de H_g por correlação, é necessário que haja uma série histórica de H_g da região, para determinar os coeficientes dos modelos empíricos, que representam características locais, o que pode inviabilizar o seu emprego em regiões onde não possui nenhuma medição (ABRAHA & SAVAGE, 2008). Os coeficientes empíricos (“a” e “b”) do modelo Angstrom-Prescott por ser uma correlação linear simples, o coeficiente linear representa a transmissividade mínima (a) onde ocorre apenas a incidência de radiação difusa, ou seja, segundo WMO (2014) abaixo dos 120 W m^{-2} não a registro nos heliogramas, e o coeficiente angular (b) que representa a radiação direta incidente, sendo seu valor influenciado pela interação da radiação com elementos da atmosfera, e a soma dos dois coeficientes (a + b) fornece a transmissividade máxima. Independente do modelo, os coeficientes de uma forma geral representam as condições climáticas local.

No contexto histórico, como descrito por Bahel et al. (1987), na década de 80 as medidas de insolação eram mais frequentes, quando comparado a da radiação global, em estações meteorológicas convencionais, porém com o passar do tempo essa relação foi se invertendo e hoje existe mais piranômetros que heliógrafos. A desvantagem dos modelos derivados da insolação é que essa variável não está

facilmente disponível, como por exemplo a temperatura do ar (ABRAHA & SAVAGE, 2008) pois os heliógrafos estão distribuídos em poucas estações meteorológicas convencionais antigas, em função da dificuldade de sua aquisição, pois são equipamentos importados, e a principal dificuldade está na parte operacional, pois necessita da presença diária de um operador para trocar as fitas e posteriormente a transcrever os valores em planilhas, porém a vantagem é que a correlação da S com a Hg é mais adequada quando comparado a temperatura do ar, como descrito por Adaramola (2012) que o desempenho do modelo baseado na insolação foi mais preciso, seguido do modelo baseado na temperatura e por último o modelo baseado na umidade relativa do ar.

Pesquisadores vem se dedicando na avaliação do desempenho dos modelos empíricos baseados na insolação em diferentes países, sendo observado que a condição climática do local é o principal fator que influencia o desempenho, como demonstrado na revisão mais abrangentes envolvendo a temática, que foi realizada por Prieto & García (2022) que encontraram um total de 165 modelos, sendo que 16 eram variações polinomiais, logarítmicos, exponenciais e potenciais com a insolação, e que foram avaliados em vários países, onde a seguir é apresentado os resultados dos artigos que abordaram a temática e avaliaram o desempenho.

Embora no geral, todos os modelos apresentaram um bom desempenho, para as 23 cidades localizadas no Iran, porém segundo Jahani et al. (2017) os modelos baseados na insolação o desempenho estatístico foi melhor quando comparado aos modelos baseados apenas na temperatura do ar e o mesmo desempenho foi observado em relação aos modelos híbridos (insolação e temperatura do ar), o que significa que a inserção da temperatura do ar no modelo não melhorou significativamente o desempenho.

Para 16 estações meteorológicas localizadas na Espanha, segundo Almorox & Hontoria (2004), o modelo empírico baseado apenas na insolação, com polinômio de terceiro grau apresentou o melhor desempenho, quando comparado aos modelos de primeiro e segundo grau e exponencial, porém, segundo os autores, todos os modelos apresentaram desempenho estatístico próximos, e em função da simplicidade os autores recomendam o emprego do modelo linear para a Espanha.

Na Turquia, de um total de 18 cidades avaliadas, apenas duas obtiveram um baixo desempenho dos modelos polinomiais de primeiro, segundo grau, logarítmico, exponencial e potencial na estima da radiação global, que segundo Bakirci (2009) é influenciado principalmente por mudança nas condições meteorológicas, relacionadas principalmente a composição da atmosfera.

O desempenho de modelos baseados na S foi superior, quando comparado aos modelos que empregam somente a T_{med} , RH_{med} e ΔT para a cidade de Iseyin, localizado na Nigéria, porém segundo Falayi et al. (2008), na falta de S pode-se utilizar outras variáveis meteorológicas, como por exemplo a temperatura do ar, sem grandes erros nas estimativas.

Em 48 cidades da China, Chen et al. (2004) encontraram que os modelos baseados na insolação, com destaque ao modelo polinomial de terceiro grau, obtiveram melhor desempenho quando comparado ao modelo baseado na temperatura do ar, e que o modelo de Bristow & Campbell não é adequado para as cidades de Xian e Chongqing pois o coeficiente “ a ” que representa transmissividade máxima para um dia claro foi de 2,35 e 3,19 respectivamente.

No estudo desenvolvido por Martim & Souza (2021), que avaliaram o desempenho do modelo de Angstrom-Prescott, em 20 estações meteorológicas, localizadas na região que compreende o bioma Amazônico brasileiro, constataram semelhanças no desempenho estatístico, nos agrupamentos sazonais ou anuais, sendo recomendado os modelos anuais com o coeficiente linear (a) oscilando de 0,2251 a 0,3561 e coeficiente angular (b) de 0,345 a 0,528 nas cidades.

Foi observado, que além da escolha do modelo empírico, a divisão da base de dados para treinamento, como por exemplo, em geral ou sazonal para treinamento dos modelos é uma das principais fases da modelagem. Estudos demonstraram que quando se compara o desempenho dos modelos baseados na S anual e sazonais não houve melhora no desempenho estatístico, como por exemplo Almorox & Hontoria (2004), que após o particionamento sazonal dos dados meteorológicos não ocorreu melhora no desempenho dos modelos na estimativa da radiação global, comparado ao anual, o mesmo resultado foi obtido por Martim & Souza (2021) com o modelo de Angstrom-Prescott avaliado na Amazônia brasileira para 20 cidades, sendo que os autores também não encontraram diferença na estimativa da radiação global com o emprego

de modelos sazonais e anuais, somente sendo observado diferença com o emprego do modelo geral com os dados das 20 estações meteorológicas, que é mais abrangente e não consegue representar as condições meteorológicas locais.

2.2.2 Modelos baseados na temperatura do ar

Os modelos empíricos que empregam a temperatura do ar são muito utilizados, pois essa é uma variável de simples monitoramento, baixo custo dos sensores, manutenção simples, está facilmente disponível e também possui uma boa correlação com a radiação global. Ao longo dos anos, vários modelos empregando a temperatura do ar foram propostos em artigos científicos e avaliados em diferentes condições climáticas.

O primeiro modelo relacionando a radiação global com a temperatura na forma de amplitude térmica ($\Delta T = T_{\max} - T_{\min}$) foi proposto por Hargreaves & Samani (1982). Já em seguida, Bristow & Campbell (1984) propuseram o modelo que possui uma correlação entre a radiação global e a função exponencial da amplitude térmica, pois segundo Goodin et al. (1999) o modelo de Bristow & Campbell, como todos os modelos baseados em temperatura é fundamentado na troca de energia entre a superfície terrestre e o ar, processo esse que aumenta a temperatura do ar, e para Abraha & Savage (2008) embora os modelos baseados na temperatura do ar possam ser considerados simples, porém, permitem estimar a Hg com baixo erro e com variáveis que estão globalmente disponíveis.

Já em 1995 com o objetivo de aumentar a eficiência de sistemas e projetos de irrigação localizados as margens do rio Senegal que percorre a África Ocidental, Hargreaves et al. (1985) propuseram o modelo baseado na ΔT local para estimar a Hg, pois era uma variável meteorológica facilmente obtida na região, e ainda avaliaram a sua influência na estimativa da evapotranspiração de referência.

Após o ano de 1995, vários artigos foram publicados avaliando os modelos já descritos acima e propondo novas metodologias, como demonstrado na revisão de literatura realizada por Pietro & García (2022) que encontraram 165 modelos, dos quais 50 modelos eram baseados apenas na temperatura do ar como variável de entrada, para vários países e condições climáticas. A seguir é apresentado os resultados de artigos que avaliaram o desempenho de modelos empíricos baseados na temperatura do ar.

Na cidade de Madinah, Arábia Saudita Benghanem & Mellit (2014) avaliaram a relação entre as variáveis meteorológicas e encontraram correlação linear de 78% entre a radiação global e a temperatura do ar média e correlação de 57% entre a radiação global e a amplitude térmica do ar (ΔT), indicando que modelos empíricos baseados na temperatura do ar apresentam um bom desempenho.

Nas regiões tropicais e subtropicais da China, Fan et al. (2018) avaliaram 20 modelos baseados na temperatura do ar e combinações híbridas, os autores relataram que embora modelos baseados somente na temperatura do ar conseguiram estimar a radiação global com boa precisão, a adição da precipitação pluviual e umidade relativa do ar melhorou consideravelmente o desempenho estatístico. Quando existe a disponibilidade apenas da temperatura do ar para a estimativa da Hg, recomenda-se o modelo com amplitude térmica (ΔT) e temperatura média do ar respectivamente.

Embora muito empregado, o modelo de Bristow & Campbell (Equação 2.33) avaliado por Goodin et al. (1999), para Manhattan obteve o menor desempenho estatístico, quando comparado ao modelo modificado proposto pelos autores, reforçando a ideia de que o desempenho dos modelos é influenciado diretamente por condições climáticas locais e também ao modelo empregado.

Na maioria dos artigos, modelos empíricos baseados na insolação apresentam desempenho superior, quando comparado a outras variáveis, porém no Egito, Hassan et al. (2016) relataram que o desempenho dos modelos calibrados e validados para a região baseados na temperatura do ar o desempenho foi superior, quando comparado ao modelo de Angstrom-Prescott encontrado na literatura para a região.

Abraha & Savage (2008) observaram que a uma relação espacial e temporal nos erros de estimativa da Hg, que é influenciado principalmente pela latitude, onde por exemplo o modelo de Bristow & Campbell superestimou os valores no inverno, primavera e no final do outono e verão e subestimou no final e início da primavera e verão. Dado que os coeficientes empíricos representam as condições meteorológicas locais, surpreendentemente, segundo Goodin et al. (1999) dois modelos com coeficientes ajustados em Manhattan e validados em 10 cidades inseridas em regiões secas e úmidas e com distâncias variando de 71 e 443 km, houve melhora no desempenho em algumas cidades, sendo que o esperado era que reduzisse o

desempenho em função da distância, sugerindo que exista outro fator fora o climático que não estava representado no modelo.

2.2.3 Modelos híbridos

A inserção de mais de uma variável meteorológica como por exemplo temperatura do ar e velocidade do vento no modelo empírico é definido como modelo híbrido, sendo que essa junção na maioria dos casos, permite melhorar significativamente a capacidade preditiva do modelo empírico. O desempenho dos modelos híbridos está diretamente relacionado com as principais variáveis meteorológicas utilizadas, como a insolação, temperatura do ar e a umidade relativa do ar (ERTEKIN & YALDIZ, 1999), associado a quantidade e a qualidade destes dados (FALAYI et al. 2008), que corrobora com os resultados obtidos por Jahani et al. (2017) em que os modelos híbridos com temperatura do ar e insolação obtiveram melhor desempenho quando comparado aos modelos simples com apenas insolação ou temperatura do ar.

No artigo de revisão de literatura, realizada por Pietro & García (2022) dos 164 modelos encontrados, 94 eram híbridos, composto por insolação, temperatura do ar, umidade relativa, latitude, pressão atmosférica, declinação solar, velocidade do vento e precipitação pluvial. Sendo assim, nos próximos parágrafos, é apresentado o panorama geral encontrado em alguns artigos que avaliaram modelos híbridos.

Na região que compreende a Bacia do rio Yangtze na China, Chen & Li (2013) concluíram que a inserção da temperatura do ar máxima e mínima (T_{max} e T_{min}) no modelo de Angstrom-Prescott, melhorou significativamente o desempenho, quando comparado o modelo de Angstrom-Prescott com a inserção da pressão de vapor d'água (VP), umidade relativa média (RHmed) e precipitação pluvial (Prec.). Para modelos baseados apenas na temperatura, a inserção de VP e RHmed melhorou significativamente o desempenho em comparação aos modelos simples, somente com uma dessas variáveis.

Na China, quando Fan et al. (2018) inseriram a precipitação pluvial (Prec.), nos modelos simples de temperatura do ar, ocorreu melhora no desempenho com destaque

para o modelo com ΔT , Prec. e Tmed e quando está disponível a RHmed, recomenda-se o modelo com ΔT , Prec., Tmed e RHmed.

No estado de Bahrain, localizado no Golfo Pérsico, Elagib et al. (1998) constataram que o desempenho do modelo híbrido com a amplitude térmica do ar (ΔT) e RHmed foi superior, quando comparado aos modelos simples, que possui apenas RHmed e Ho.

O desempenho do modelo de Angstrom-Prescott simples, quando comparado a modelos híbridos com a inserção de parâmetros como a temperatura média do ar e umidade relativa média, para a cidade de Jeddah, Arábia Saudita, segundo El-Sebaai et al. (2009) melhorou o desempenho na estimativa da radiação global.

Modelos simples com apenas uma variável meteorológica (S, Tmed, RHmed e ΔT) segundo Falayi et al. (2008) obteve menor desempenho com R^2 de 0,5669 a 0,8746, MBE de -0,00018 a -0,00461 MJ m² d⁻¹ e RMSE de 0,037 a 0,069 MJ m² d⁻¹, quando comparado aos modelos híbridos com duas, três e quatro variáveis de entrada com R^2 de 0,8984 a 0,9748, MBE de -0,00019 a 0,00026 MJ m² d⁻¹ e RMSE de 0,017 a 0,034 MJ m² d⁻¹ respectivamente.

2.2.4 Modelos baseados em variáveis astronômicas

Os modelos com base em variáveis astronômicas permitem estimar a Hg sem o registro de dados meteorológicos, como por exemplo a declinação solar, dia juliano e radiação extraterrestre, porém são estimativas aproximadas (PIETRO & GARCÍA, 2022), pois não considera as condições atmosféricas que interferem diretamente na transmissividade da Hg, sendo assim, são modelos limitados, porém devem ser avaliados para cada condição.

Vários trabalhos vêm se dedicando na avaliação do desempenho dos modelos empíricos baseados em variáveis astronômicas, como por exemplo para a cidade de Madinah, localizada na Arábia Saudita, Benghanem & Mellit (2014) encontraram uma correlação de 90% entre a radiação global e a radiação extraterrestre.

Para a cidade de Antalya na Turquia, Ertekin & Yaldiz (1999) encontraram uma correlação de 0,984 para a relação entre a Hg e a δ .

Em Amman na Jordânia, Al-Salaymeh (2006) utilizando a função trigonométrica onda senoidal com o dia juliano (DJ), correlação de Lorentzin, correlação Gaussiana e polinomial de quarto grau, obteve uma correlação de 0,96 para todos os modelos, porém segundo os autores a radiação global e a insolação sofrem influência dos elementos atmosféricos, causando espalhamento que se intensifica principalmente nos meses de janeiro e dezembro. Dependendo das condições atmosféricas, podem ser inseridas variáveis com melhor correlação com Hg nos modelos baseados em variáveis astronômicas como por exemplo a insolação.

2.3 INTELIGÊNCIA ARTIFICIAL (IA) E MACHINE LEARNING (ML)

A inteligência artificial (IA) é uma grande área do conhecimento no campo da ciência da computação, matemática e estatística, que é subdividida em várias subáreas (Figura 2.3), entre elas está o machine learning ou traduzido para o português aprendizado de máquina (ML), que se refere a uma série de técnicas, que através do processo conhecido como treinamento consegue encontrar padrões em um determinado conjunto de dados e assim prever valores futuros. Existem várias técnicas de ML, entre os mais conhecidos e utilizados estão a rede neural artificial (ANN) e suporte de vetor e máquina (SVM) (Agbulut et al. 2021).



Figura 2.3. Inteligência artificial, aprendizado de máquina, aprendizado supervisionado, regressão, classificação (Adaptado de Agbulut et al. 2021).

As técnicas de machine learning ao longo dos anos vem sendo avaliados e população para a solução de diversos problemas do cotidiano (SANTOS et al. 2021), o que permitiu um desenvolvimento acelerado e aplicações em diversas áreas do conhecimento (ZHOU et al. 2021).

2.3.1 Rede Neural Artificial (ANN)

A rede neural artificial (ANN) foi desenvolvida a partir da observação da estrutura da rede neural biológica, sendo que no início o desenvolvimento da pesquisa foi lento, pois havia a necessidade de computadores com boa capacidade de processamento. Atualmente, as redes neurais artificiais são empregadas em vários setores da sociedade, como na resolução de problemas com modelagem linear ou não linear, problemas com centenas ou milhares de variáveis de entrada e que necessitam de uma solução rápida.

O neurônio é a estrutura básica de uma ANN e possui capacidade de processamento de informações de forma individual (AGBULUT et al. 2021) onde cada neurônio é responsável pela operação de soma da multiplicação do sinal de entrada (X) pelo peso (W) de cada informação, fornecendo o sinal de saída (U) do neurônio (QUEI et al., 2017) conforme a Equação 2.65.

$$U = (\sum_{i=1}^n w * x + \beta a) \quad 2.65$$

A estrutura básica de uma rede neural artificial do tipo multi-layer perceptron (MLP) pode ser representada pela Figura 2.4, com um sistema interligado através de conexões ponderadas que é composto por variáveis de entrada (X) que estão na camada de entrada, camada oculta (CO) e camada de saída (CS) e são interligados por pesos (W) e que matematicamente pode ser expressa pela Equação 2.66 (FAN et al. 2019).

$$R_s = f(\sum_{j=1}^n w * x + \beta a) \quad 2.66$$

$$f(u) = \frac{1}{[1+\exp(-u)]} \quad 2.67$$

As variáveis de entradas são informações que possuem correlação com a saída da rede, e geralmente essas variáveis são normalizadas com o objetivo de minimizar a influência que as diferentes unidades de medidas possam ter no treinamento.

Os pesos (W) são elementos de ligação entre as diferentes camadas de uma rede neural artificial, que fornece um valor numérico com a importância de variável de entrada na predição da variável de saída (ZHOU et al. 2021).

Em boa parte dos problemas a relação entre os dados de entrada e saída não é direta, pois não passa pela origem, sendo necessário adicionar uma constante conhecida como bias ou viés (β) que é uma forma fácil de corrigir esta distorção, pois com a inserção do β a rede construída melhora a capacidade de identificar padrões mais complexos nos dados e melhora a capacidade de generalização.

A função de ativação ($f(U)$) permite que uma grande mudança nos valores de entrada, pesos ou bias cause uma pequena mudança na camada de saída, melhorando a capacidade preditiva da ANN. Existem várias funções de ativação, entre elas a mais utilizada é a função sigmoide (Equação 2.67) e dependendo do problema pode ser empregado a função constante, linear, gaussiana, softmax, tangente hiperbólica, ReLU.

Cada camada pode conter vários neurônios, podendo ser definido como elemento de ligação entre as diferentes camadas e a menor unidade de processamento dentro da ANN. Toda a modelagem da ANN pode ser expressa pela Equação 2.66, com a soma da multiplicação das variáveis de entrada pelo seu respectivo peso e a soma do viés. O número de neurônio na camada oculta pode variar de acordo com o problema a ser resolvido, porém no geral o número deve ser igual ou próximo a dimensionalidade dos dados de entrada (BELLIDO-JIMÉNEZ et al. 2021).

O número de camadas que está entre a camada de entrada e a camada de saída é definida como camada oculta (CO), sendo que para a maioria dos problemas simples que são linearmente separáveis, uma camada oculta pode ser o suficiente, enquanto para problemas mais complexos pode ser necessário redes mais profundas conhecidas como Deep Learning com mais de uma camada oculta.

Quando se trabalha com a ANN, alguns hiperparâmetros são essenciais para um bom desempenho da rede e devem ser definidos antes do treinamento, que são eles a taxa de aprendizagem, momentum, número de interações, número de camadas ocultas e número de neurônios. Taxa de aprendizagem (Learning Rate) controla o ritmo de atualização dos pesos, quando muito alto pode oscilar em torno de um mínimo global, já quando o valor é baixo o treinamento pode ser muito lento ou convergir em

torno de mínimos locais (SILVA et al. 2017). O momentum visa melhorar a convergência e evitar oscilações excessivas durante o treinamento em direção ao mínimo global (ZHOU et al. 2021). O número de interações é o número de vezes que todo o conjunto de treinamento é passado pela rede até que o erro se estabilize.

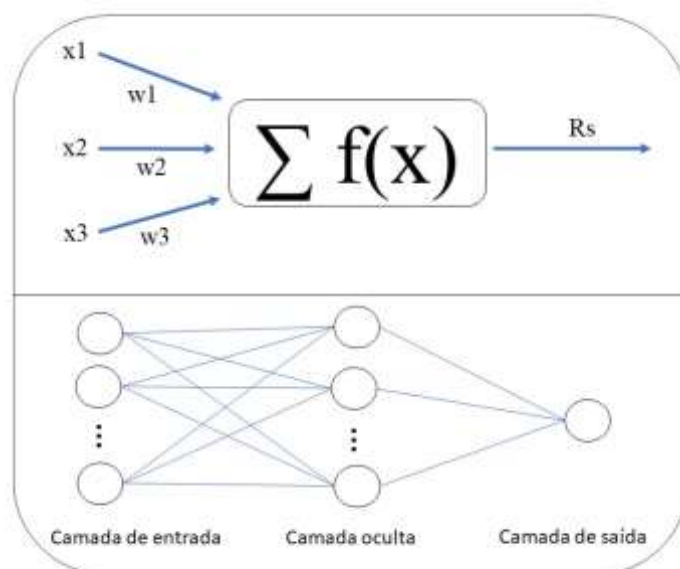


Figura 2.4. Estrutura básica de uma rede neural artificial (RNA) (Adaptado de Santos et al. 2021).

A técnica mais utilizada para treinamento da ANN é retropropagação do erro (backpropagation), que consiste em um processo iterativo de determinação do erro e ajuste dos pesos da rede (SILVA et al. 2017) e funciona da seguinte forma, no feedforward o treinamento segue o sentido direto com as variáveis de entrada sendo multiplicadas pelos pesos, como sintetizado na Equação 2.66, sendo encontrado o valor de saída que permite a determinação do erro na estimativa, já no sentido inverso do processo conhecido como feedback o erro é utilizado para a atualização dos pesos no sentido da camada de saída para a camada de entrada, sendo que esse processo ocorre iterativamente até que o erro encontrado seja o mínimo possível.

2.3.2 Suporte de vetor e máquina (SVM)

A máquina de vetores de suporte (SVM) é uma técnica de aprendizado de máquina supervisionado usado para tarefas de classificação e regressão (QUEJ et al. 2017), sendo muito eficiente em problemas de alta dimensão e com dados que não são

linearmente separáveis (VAPNIK, 1995), além disso, a dimensão é representada pelo número de variáveis de entrada (BELLIDO-JIMÉNEZ et al. 2021).

O SVM tanto de classificação quanto de regressão pode ser exemplificado pela Figura 2.5, onde a linha central é conhecida como hiperplano, a distância entre o hiperplano e a linha superior ou linha inferior é conhecido como margem. Na classificação o objetivo do hiperplano é ficar o mais distante possível das classes, já na regressão o objetivo muda e a função do hiperplano é ficar o mais próximo possível dos pontos. Os vetores de suporte são os pontos de dados mais próximos da fronteira de decisão (hiperplano) e tem influência sobre o posicionamento da fronteira.

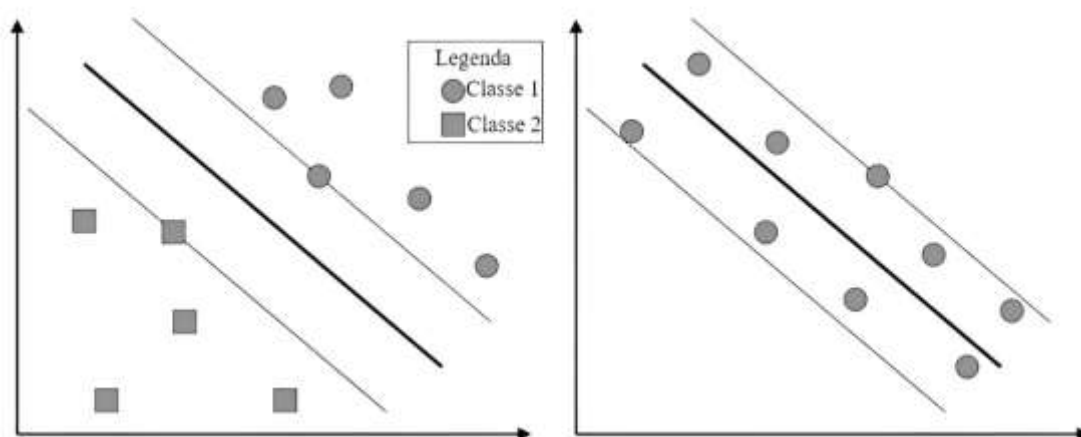


Figura 2.5. SVM de a) classificação (classe 1 e 2) e b) regressão (Adaptado de Bellido-Jiménez et al. 2021).

Existem várias funções de Kernel empregadas no SVM, como linear (Equação 2.69), polinomial (Equação 2.70) com $\gamma > 0$, sigmoide (Equação 2.71) e de base radial gaussiana (Equação 2.72) com $\gamma > 0$ (Quej et al. 2017), sendo que a função de base radial de Kernel (BRF) gaussiana é a mais utilizada nos diferentes problemas pois é considerada a mais eficiente, simples e capaz de resolver os problemas complexos (SANTOS et al. 2021) pois estima uma série de regressões simples (Equação 2.68) com base em várias funções de Kernel (Equação 2.72), ou seja, transforma os dados originais de alta dimensão em dados de baixa dimensão facilitando a separação linear dos dados (FAN et al. 2019).

$$K(x_1, x_n) = \varepsilon K(x_1, x_n) + b \quad 2.68$$

$$K(x_1, x_n) = x_1 * x_n \quad 2.69$$

$$K(x_1, x_n) = (\gamma x_1 * x_n + r)^d \quad 2.70$$

$$K(x_1, x_n) = \tanh(\gamma x_1 * x_n + r) \quad 2.71$$

$$K(x_1, x_n) = \exp(-\gamma \|x_1 - x_n\|^2) \quad 2.72$$

Onde o $K(x_1, x_n)$ é a função de redução da dimensão de entrada do vetor (x_1, x_n) , ε e b são vetores de peso e polarização, $(x_1 - x_n)^2$ é a distância euclidiana entre os valores, e o d , r e γ são os parâmetros fornecidos a função de Kernel.

Alguns parâmetros como custo (C), gama (γ) e épsilon (ε) devem ser fornecidos, porém segundo Quej et al. (2017) a escolha destes valores deve atender as características do conjunto de dados, sendo determinados através de técnicas de otimização dos hiperparâmetros ou validação cruzada, pois um valor maior de C implica em uma penalização maior por erros no conjunto de treinamento, levando a busca por uma margem menor, com valores mais precisos no treinamento e com menor capacidade de generalização (maior erro na validação), porém quando o valor de C é reduzido ocorre a suavização da margem, permitindo uma classificação com menor rigor no treinamento. O parâmetro γ está associado ao Kernel utilizado, onde o valor baixo implica em fronteiras de decisões mais suaves (mais lineares), já o valor alto leva as fronteiras de decisão são mais complexas, podendo se adaptar melhor a dados não lineares. O parâmetro ε controla o quanto a margem pode ser extrapolada no treinamento, se for zero todos os pontos devem estar corretamente classificados e estarão na margem ou no lado correto do hiperplano de decisão, sendo que o ε permite uma certa flexibilidade na classificação para lidar com casos em que os dados não são linearmente separáveis ou possuem outliers.

2.3.3 Estimativa da radiação global com MLP e SVM

A radiação global (Hg) é o termo que se refere a quantidade total de energia proveniente do sol incidente na superfície terrestre, sendo essa a principal fonte de energia na Terra, sendo também uma variável com aplicação em diversas áreas do conhecimento, porém a sua medição está restrita a poucos locais, pois necessita de estações meteorológicas solarimétricas com custo ainda elevado (SILVA et al. 2021).

A radiação global possui boa correlação com outras variáveis que estão disponíveis como insolação, temperatura do ar, umidade relativa do ar, velocidade do vento e em algumas regiões até variáveis astronômicas como radiação no topo da atmosfera e fotoperíodo, porém essa correlação possui uma dependência direta com as

condições atmosféricas locais, sendo necessário avaliar separadamente para cada região, com isso a partir de 2003 o número de pesquisas que vêm sendo desenvolvidas para avaliar o desempenho das diferentes técnicas de machine learning na estimativa da radiação global aumentou exponencialmente (ZHOU et al. 2021).

Além disso, trabalhos acadêmicos vem se dedicando em avaliar as diferentes combinações de entrada (SANTOS et al. 2021, SILVA et al. 2017 e BELLIDO-JIMÉNEZ et al. 2021), como insolação, temperatura do ar, umidade relativa do ar, variáveis astronômicas como radiação extraterrestre entre outros (ZHOU et al. 2021) sendo que a temperatura é a variável de fácil monitoramento (SANTOS et al. 2021), porém independente da técnica utilizado e da combinação das variáveis de entrada, a necessidade de treinamento e validação com dados locais, para avaliar o desempenho.

Em todos os trabalhos que utilizaram o SVM na estimativa da radiação global, quando empregado a função de base radial de Kernel (BRF) gaussiana, deve ser fornecido os parâmetros como C , γ e ϵ . Como observado na Tabela 2.1, dependendo da região avaliada, e das variáveis empregadas, os parâmetros podem mudar, com o C de 0 a 512, γ de 0,046 a 300 e ϵ de 0,001 a 1000000, sendo que o melhor hiperparâmetro pode mudar, de acordo com as variáveis de entrada.

Tabela 2.1. Parâmetros C , γ e ϵ utilizados no SVM para estimativa da radiação global.

Literatura	Parâmetros SVM			Função de ativação
	C	γ	ϵ	
Santos et al. (2021)	100	0,3	0,001	RBF Kernel
Silva et al. (2017)	100	0,3	0,001	RBF Kernel
Agbulut et al. (2021)	0 a 200	300	1000000	RBF Kernel
Fan et al. (2019)	40 a 60	20 a 40	-	RBF Kernel
Quej et al. (2017)	1,08 a 9,74	0,046 a 0,702	0,287 a 0,533	RBF Kernel
Ghimire et al. (2019)	32 a 512	0,125 a 0,625	0,01	RBF Kernel

Nos estudos que utilizaram a rede neural artificial (ANN) do tipo MLP a taxa de aprendizado (TA) foi de 0,2 a 0,3, o momentum (MO) foi de 0,1 a 0,2, número de interações (IT) foi de 500 a 1000, e as variáveis da camada de entrada (VCE) em alguns trabalhos foram avaliadas diferentes combinações de entrada variando de 2 a 6 ou valor fixo de variáveis de 5 a 20, o número de neurônios na camada oculta foi de 1 a 52 e a função de ativação (FA) mais utilizada é a sigmoide (Tabela 2.2), demonstrando que

o valor dos melhores hiperparâmetros pode mudar, de acordo com as variáveis de entrada escolhidas.

Tabela 2.2. Parâmetros dos modelos de rede neural artificial (ANN).

literatura	TA	MO	IT	VCE	neurônios	FA
Santos et al. (2021)	0,2	0,1	1000	2 a 5	-	sigmoide
Silva et al. (2017)	0,3	0,2	500	2 a 6	1 a 10	sigmoide
Agbulut et al. (2021)	-	-	-	5	8 a 16	-
Bounoua et al. 2021	-	-	-	13	10	-
Fan et al. (2019)	-	-	-	-	5 a 7	-
Quej et al. (2017)	-	-	-	3 e 4	12 a 52	sigmoide
Ghimire et al. (2019)	-	-	-	20	13 a 28	linear, logística e tangente

Nos estudos que utilizaram somente uma combinação de entrada (Tabela 2.3) as variáveis de entrada são dia juliano (DJ), temperatura máxima, média, mínima (Tmax, Tmed e Tmin), velocidade do vento (Vv), direção média diária (Dir), radiação do topo da atmosfera (Ho), pressão de vapor de água (Pvap), amplitude térmica ($\Delta T = T_{max} - T_{min}$), raiz quadrada da amplitude térmica ($\sqrt{\Delta T}$), razão da temperatura (Tmin/Tmax), nebulosidade, fotoperíodo (So), dados de poluição do ar como AQI que descreve quantitativamente a qualidade do ar, PM_{2,5} que são partículas finas, o PM₁₀ são partículas inaláveis e O₃ o ozônio.

Tabela 2.3. Combinações de entrada utilizada em vários trabalhos com técnicas de ML.

Referências	Variáveis de entrada
Agbulut et al. (2021)	Tmax, Tmin, nebulosidade, Ho e So DJ, Tmed, Rhmed, Vv, Dir, Pvap, Ho, Tmax, Tmin, ΔT , $\sqrt{\Delta T}$ e Tmin/Tmax
Bounoua et al. 2021	Tmin/Tmax
Jia et al. (2022)	S, So, AQI, PM _{2,5} , PM ₁₀ e O ₃
Quej et al. (2017)	Tmax, Tmin, Prec. e Ho Tmin, Tmax e Ho

Na literatura existem vários trabalhos que se propuseram a avaliar diferentes técnicas de machine learning para estimativas de radiação global, sendo que dentre os principais estão a rede neural artificial (ANN) e máquina de vetores e suporte (SVM) e diferentes combinações de entrada em diversos países entre eles estão a Turquia (AGBULUT et al. 2021), Marrocos (BOUNOUA et al. 2021), Espanha e Estados Unidos (BELLIDO-JIMÉNEZ et al. 2021), China (FAN et al. 2019; HE et al. 2020),

México (QUEJ et al. 2017) com classificação climática Aw, Austrália (GHIMIRE et al. 2019) e no Brasil para a cidade de Campo Grande (SANTOS et al. 2021) com classificação climática Aw e Botucatu (SILVA et al. 2017) com classificação climática Cwa, porém percebe-se para as condições brasileiras os estudos são locais com pequena abrangência espacial.

Para a cidade de Campo Grande, capital do Estado de Mato Grosso do Sul, Santos et al. (2021) avaliaram seis diferentes combinações de entrada (Tabela 2.4) com a temperatura do ar (Tmax, Tmed e Tmin) e radiação no topo da atmosfera (Ho) com as técnicas de rede neural artificial (ANN1 a ANN6) e suporte de vetor e máquina (SVM1 a SVM6) na estimativa da radiação global (Figura 2.10).

Tabela 2.4. Combinações de entrada para ANN e SVM (Adaptado de Santos et al. 2021).

Modelo		Variáveis de entrada
ANN1	SVM1	Ho, Tmax
ANN2	SVM2	Ho, Tmin
ANN3	SVM3	Ho, ΔT
ANN4	SVM4	Ho, Tmed
ANN5	SVM5	Ho, Tmax e Tmin
ANN6	SVM6	Ho, Tmax, Tmin, ΔT e Tmed

Na validação dos modelos ANN e SVM segundo Santos et al. (2021) a combinação com as variáveis Ho, Tmin apresentou o pior desempenho, seguido da combinação Ho, Tmed (ANN4 e SVM4), já o melhor desempenho foi obtido com as variáveis Ho, Tmax, Tmin, ΔT e Tmed (ANN6 e SVM6). Na comparação entre as duas técnicas o desempenho do SVM foi melhor quando comparado a ANN em todas as combinações avaliadas.

Silva et al. (2017) avaliaram quatro combinações de entrada, com as técnicas MLP e SVM para a cidade de Botucatu, sendo o primeiro com as variáveis do modelo de A-P com Ho e S/So (ANN1 e SVM1), a segunda adicionando a temperatura do ar com Ho, Tmax, Tmin e S/So (ANN2 e SVM2), a terceira combinação adicionaram a precipitação pluvial com Ho, Tmax, Tmin, Prec. e S/So (ANN3 e SVM3) e a quarta adicionaram todas as variáveis com Ho, Tmax, Tmin, Prec., RHmed e S/So (ANN4 e SVM4) na estimativa da radiação global com ANN e SVM para anos típicos e atípicos. Segundo os autores, os resultados mostraram que embora em todas as combinações o

desempenho foi bom, porém, a melhor combinação foi a ANN4 com r de 0,95 a 0,96 MBE de -2,30 a 1,38 MJ m⁻² d⁻¹, RMSE de 2,68 a 2,89 MJ m⁻² d⁻¹ e SVM4 com r de 0,960 a 0,9700, MBE de -0,47 a 0,28 MJ m⁻² d⁻¹ e RMSE de 1,64 a 2,14 MJ m⁻² e o desempenho da técnica SVM foi superior quando comparado ao ANN em todas as combinações avaliadas.

Como o resultado do estudo desenvolvido por Bounoua et al. (2021), o desempenho das técnicas de ML na estimativa da radiação global sofreu interferência das condições microclimáticas, com o melhor desempenho sendo obtido na cidade de Zagora com coeficiente de correlação de 0,9469 a 0,9619, porém a cidade costeira de Tan-Tan, foi encontrado o pior desempenho, com correlação de 0,8173 a 0,8753. Quando se compara as técnicas, a ANN obteve o pior desempenho, já o Bagging e RF o melhor desempenho.

Na avaliação de 12 modelos de ML em 50 cidades e 4 zonas climáticas, sendo a zona continental temperada (TCZ), zona de monção temperada (TMZ), zona de planalto montanhoso (MPZ) e zona de monção subtropical (SMZ) na China, no estudo desenvolvido por Fan et al. (2019), encontraram (Figura 2.13) que todos os modelos de ML exceto o RF, M5Tree e GRNN foram capazes de estimar a radiação global com grande precisão em todas as zonas climáticas. O desempenho do modelo SVM para as cidades com o MBE foi de -0,437 a 0,637 MJ m⁻² d⁻¹ na zona climática TCZ e SMZ, o RMSE foi de 2,07 a 2,53 MJ m⁻² d⁻¹ (TCZ e SMZ) e R² foi de 0,880 a 0,933 (SMZ e TCZ), já o MLP foi obtido o MBE de -0,284 a 0,612 MJ m⁻² d⁻¹ (TCZ e SMZ), RMSE foi de 2,06 a 2,57 MJ m⁻² d⁻¹ (TCZ e SMZ) e R² foi de 0,880 a 0,934 (SMZ e TCZ).

He et al. (2020) avaliaram diferentes combinações para estimar a radiação global em 80 cidades da China, com as diferentes combinações da S, Ho, Tmax, Tmed, Tmin, So, Prec., Pt, ΔT e RHmed e Vv com as combinações simples e híbridas. O SVM quando empregado somente a So e Ho como variável de entrada, o desempenho estatístico foi o pior, pois a radiação global na região sofre grande influência da precipitação pluvial. Nas combinações híbridas com a temperatura do ar e precipitação pluvial, o pior desempenho com o SVM avaliando o RMSE foi obtido com as combinações Ho + Prec e Ho + Pt, seguido das combinações Ho + ΔT + Prec., Ho + ΔT + Pt e o melhor desempenho para todas as cidades da China foi a combinação S + Ho + Prec. ou S + Ho + Pt. Já as melhores combinações são S + Ho + Tmax + Tmed

+ Tmin + RHmed, S + Ho + ΔT + Tmax + RHmed + Pt + Tmed + Vv, S + Ho + Tmax + Tmed + Tmin + RHmed e So + Ho + Tmax + Tmed + Tmin + RHmed + Pt nas zonas climáticas MPZ, SMZ, TCZ e TMZ. Conforme a análise da métrica estatística RMSE para estimativa da radiação global na China, recomenda-se utilizar o SVM.

Foi avaliado as técnicas MLP, ANFIS e SVM na estimativa da radiação global com duas combinações de entrada, sendo a primeira com Tmax, Tmin, Prec., Ho e a segunda com Tmax, Tmin e Ho. Segundo Quej et al. (2017) o melhor desempenho foi obtido empregando o SVM e as variáveis Tmax, Tmin, Prec. e Ho, sendo que a inserção da precipitação pluvial melhorou a capacidade preditiva da radiação global. O melhor desempenho do SVM foi obtido na cidade de Efraín Hernández com R^2 de 0,732, RMSE de $2,69 \text{ MJ m}^{-2} \text{ d}^{-1}$ e MAE de $2,10 \text{ MJ m}^{-2} \text{ d}^{-1}$ e o pior na cidade de Celestún com R^2 de 0,635, RMSE de $2,82 \text{ MJ m}^{-2} \text{ d}^{-1}$ e MAE de $2,16 \text{ MJ m}^{-2} \text{ d}^{-1}$.

Em cinco cidades localizadas na Austrália, das 85 variáveis disponíveis segundo Ghimire et al. (2019) após serem submetidas a análise da técnica análise de componentes de vizinhança (NCA), foram separadas as 20 melhores, que foram utilizadas para a estimativa da radiação global. No gráfico de dispersão entre a radiação global medida e estimada, o melhor desempenho foi obtido com o MLP com correlação de 0,939 a 0,967 para as cidades de Gold Coast e Brisbane e com o SVR a correlação foi de 0,921 a 0,959 para as cidades de Sunshine Coast e Toowoomba.

Em quatro cidades da Turquia, Agbulut et al. (2021) avaliaram quatro técnicas de ML, sendo que o desempenho do MLP foi o R^2 de 0,8830 a 0,9360, MBE de -0,0251 a 0,231 MBE de -0,0251 a 0,231 MJ m^{-2} e RMSE de 2,15 a 2,44 MJ m^{-2} e SVM com R^2 de 0,8800 a 0,9260, MBE de -0,0678 a 0,289 MJ m^{-2} e RMSE de 2,27 a 2,82 MJ m^{-2} , sendo que nesse estudo o desempenho do MLP foi superior quando comparado ao SVM.

Para 8 cidades da China, Jia et al. (2022) avaliaram a estimativa da radiação global com vários modelos de ML, sendo eles o SVM, RF e GLMNET com S, So e dados de poluição do ar, sendo que o desempenho do SVM foi melhor com correlação de 0,52 a 0,98, seguido do modelo GLMNET 0,55 a 0,96 e o pior desempenho foi encontrado com o modelo RF com correlação de 0,24 a 0,91.

2.4 REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHA, M. G. SAVAGE, M. J. Comparasion of estimates of daily solar radiation from air temperature range for application in crop simulations. **Agriculture and Forest Meteorology**, v. 148, p. 401-416, 2008. <https://doi.org/10.1016/j.agrformet.2007.10.001>.

ADARAMOLA, M. S. Estimating global solar radiation using common meteorological data in Akure, Nigeria. **Renewable Energy**, v. 47, p. 38-44, 2012. <https://doi.org/10.1016/j.renene.2012.04.005>.

ALMOROX, J. HONTORIA, C. Global solar radiation estimation using sunshine duration in Spain. **Energy Conversion and Management**, v. 45, p. 1529-1535, 2004. <https://doi.org/10.1016/j.enconman.2003.08.022>.

ALVARES, C. A. STAPE, J. L. SENTELHAS, P. C. GONÇALVES, J. L. M. SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013. DOI: 10.1127/0941-2948/2013/0507.

AL-SALAYMEH, A. Modeling of global daily solar radiation on horizontal surfaces for amman city. **Emirates Journalfor Engineerring Research**, v. 11, p. 49-56, 2006.

ANGSTROM, A. Solar and terrestrial radiation. **Quarterly Journal of the Royal Meteorological Society**, v. 50, p. 121-126, 1924. <https://dx.doi.org/10.1002/qj.49705021008>.

AGBULUT, Ü. GÜREL, A. E. BIÇEN, Y. Prediction of daily global solar radiation using diferente machine learning algorithms: Evaluation and comparation. **Renewable and Sustainable Energy Reviews**, v. 135, 2021. <https://doi.org/10.1016/j.rser.2020.110114>.

BAKIRCI, K. Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. **Energy**, v. 34, p. 485-501, 2009. <https://doi.org/10.1016/j.energy.2009.02.005>.

BAHEL, V. BAKHSH, H. SRINIVASAN, R. A correlation for estimation of global solar radiation. **Energy**, v. 12, p. 131-135, 1987. [https://doi.org/10.1016/0360-5442\(87\)90117-4](https://doi.org/10.1016/0360-5442(87)90117-4).

BENGHANEM, M. MELLIT, A. A simplified calibrated model for estimating daily global solar radiation in Madinah, Saudi Arabia. **Theoretical and Aplied Climatology**, v. 115, p. 197-205, 2014. DOI: 10.1007/s00704-013-0884-2.

BRISTOW, K. L.; CAMPBELL, G. S. On the relationship between incoming solar radiation and daily maximum and minimum temperature. **Agricultural and Forest Meteorology**, v. 31, n. 2, p. 159-166, 1984. [https://doi.org/10.1016/0168-1923\(84\)90017-0](https://doi.org/10.1016/0168-1923(84)90017-0).

BELLIDO-JIMÉNEZ, J. A. GUALDA, J. E. GARCÍA-MARÍN, A. P. Assessing new intra-daily temperature-based machine learning models to outperform solar radiation predictions in diferente conditions. **Applied Energy**, v. 298, 2021. <https://doi.org/10.1016/j.apenergy.2021.117211>.

- BOUNOUA, Z. CHAHIDI, L. O. MECHAQRANE, A. Estimation of daily global radiation using empirical and machine-learning methods: A case study of five Moroccan locations. **Sustainable Materials and Technologies**, v. 28, 2021. <https://doi.org/10.1016/j.susmat.2021.e00261>.
- CHEN, R. ERSI, K. YANG, J. LU, S. ZHAO, W. Validation of five global radiation models with measured daily data in China. **Energy Conversion and Management**, v. 45, p. 1759-1769, 2004. <https://doi.org/10.1016/j.enconman.2003.09.019>.
- CHEN, J.-L.; LI, G.-S. Estimation of Monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. **International Journal of Climatology**, v. 33, n. 2, p. 487-498, 2013. DOI: 10.1002/joc.3442.
- ERTEKIN, C. YALDIZ, O. Estimation of monthly average daily global radiation on horizontal surface for Antalya (Turkey). **Renewable Energy**, v. 17, p. 95-102, 1999. [https://doi.org/10.1016/S0960-1481\(98\)00109-8](https://doi.org/10.1016/S0960-1481(98)00109-8).
- ELAGIB, N. A. BABIKER, S. F. ALVI, S. H. New empirical models for global solar radiation over Bahrain. **Energy Conversion and Management**, v. 39, p. 827-835, 1998. DOI: 10.1016/S0196-8904(97)00035-6.
- EL-SEBAILI, A. A. AL-GHAMDI, A. A. AL-HAZMI, F. S. FAIDAH, A. S. Estimation of global solar radiation on horizontal surfaces in Jeddah, Saudi Arabia. **Energy Policy**, v. 37, p. 3645-3649, 2009. <https://doi.org/10.1016/j.enpol.2009.04.038>.
- FAN, J. WU, L. ZHANG, F. CAI, H. ZENG, W. WANG, X. ZOU, H. Empirical and machine learning models for predicting daily global solar radiation from Sunshine duration: A review and case study in China. **Renewable and Sustainable Energy Reviews**, v. 100, p. 186-212, 2019. <https://doi.org/10.1016/j.rser.2018.10.018>.
- FALAYI, E. O. ADEPITAN, J. O. RABIU, A. B. Empirical models for the correlation of global solar radiation with meteorological data for Iseyin, Nigeria. **International Journal of Physical Sciences**, p. 3, v. 210-216, 2008.
- FAN, J. CHEN, B. WU, L. ZHANG, F. LU, X. XIANG, Y. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. **Energy**, v. 144, p. 903-914, 2018. <https://doi.org/10.1016/j.energy.2017.12.091>.
- GOODIN, D. G. HUTCHIMSON, J. M. S. VANDERLIP, R. L. KNAPP, M. C. Estimating solar irradiance for crop modeling using daily air temperature data. **Agronomy Journal**, v. 91, p. 845-851, 1999. <https://doi.org/10.2134/agronj1999.915845x>.
- GHIMIRE, S. DEO, R. C. DOWNS, N. J. RAJ, N. Global solar radiation prediction by ANN integrated with European Centre for medium range Weather forecast fields in solar rich cities of Queensland Australia. **Journal of Cleaner Production**, v. 216, n. 10, p. 288-310, 2019. <https://doi.org/10.1016/j.jclepro.2019.01.158>.
- HARGREAVES, G. L. ASCE, A. M. HARGREAVES, G. H. ASCE, F. ASCE, RILEY, J. P. Irrigation water requirements for Senegal river basin. **Journal of**

Irrigation and Drainage Engineering, v. 111, p. 265-275, 1985. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1985\)111:3\(265\)](https://doi.org/10.1061/(ASCE)0733-9437(1985)111:3(265)).

HASSAN, G. E. YOUSSEF, M. E. MOHAMED, Z. E. ALI, M. A. HANAFY, A. A. New Temperature-based Models for Predicting Global Radiation. **Applied Energy**, v. 179, n. 1, p. 437-450, 2016. <https://doi.org/10.1016/j.apenergy.2016.07.006>.

HE, C. LIU, J. XU, F. ZHANG, T. CHEN, S. SUN, Z. ZHENG, W. WANG, R. HE, L. FENG, H. YU, Q. HE, J. Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods. **Energy Conversion and Management**, v. 220, n. 15, 2020. <https://doi.org/10.1016/j.enconman.2020.113111>.

JAHANI, B. DINPASHOH, Y. NAFCHI, A. R. Evaluation and development of empirical models for estimating daily solar radiation. **Renewable and Sustainable Energy Reviews**, n. 73, p. 878-891, 2017. <https://doi.org/10.1016/j.rser.2017.01.124>.

JIA, D. YANG, L. LV, T. LIU, W. GAO, X. ZHOU, J. Evaluation of machine learning models for predicting daily global and diffuse solar radiation under different Weather/pollution conditions. **Renewable Energy**, v. 187, p. 896-906, 2022. <https://doi.org/10.1016/j.renene.2022.02.002>.

PRIETO, JESUS-IGNACIO, GARCÍA, D. Global solar radiation models: A critical review from the point of view of homogeneity and case study. **Renewable and Sustainable Energy Reviews**, v. 155, 2022. <https://doi.org/10.1016/j.rser.2021.111856>.

PRESCOTT, J. Evaporation from a Water Surface in Relation to Solar Radiation. **Transactions of the Royal Society of South Australia**, v. 64, p. 114-118, 1940.

SANTOS, C. M. TERAMOTO, E. T. SOUZA, A. ARISTONE, F. IHADDADENE, R. Several models to estimate daily global solar irradiation: adjustment and evaluation. **Arabian Journal of Geosciences**, v. 14, n. 4, 2021. <https://doi.org/10.1007/s12517-021-06603-8>.

SILVA, M. B. P.; ESCOBEDO, J. F.; ROSSI, T. J.; SANTOS, C. M.; SILVA, S. H. M. G. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 160, p. 11-23, 2017. <https://doi.org/10.1016/j.jastp.2017.04.001>.

SOUZA, A. P. SILVA, A. C. TANAKA, A. A. ULIANA, E. M. ALMEIDA, F. T. KLAR, A. E. GOMES, A. W. A. Global radiation by simplified models for the state of Mato Grosso, Brazil. **Pesquisa Agropecuária Brasileira**, v. 52, n. 4, p. 215-227, 2017. <https://doi.org/10.1590/S0100-204X2017000400001>.

MARTIM, C. C. ZAMADEI, T. SOUZA, A. P. ALMEIDA, F. T. ZOLIN, C. A. COEFICIENTES DE ANGSTRÖM-PRESCOTT E A EVAPOTRANSPIRAÇÃO DE REFERÊNCIA NA TRANSIÇÃO CERRADO-AMAZÔNIA DO MATO GROSSO. **Revista Brasileira de Climatologia**, v. 26, p. 579-594, 2020. <https://10.5380/abclima.v26i0.62369>.

MARTIM, C. C. SOUZA, A. P. Estimativas da radiação global com base na insolação na Amazônia brasileira. **Revista Ibero-Americana de Ciências Ambientais**, v. 12, n. 10, p. 233-246, 2021. <https://doi.org/10.6008/CBPC2179-6858.2021.010.0020>.

QUEJ, V. H. ALMOROX, J. ARNALDO, J. A. SAITO, L. ANFIS, SVM e ANN soft-computing techniques to estimated daily global solar radiation in a warm sub-humid environment. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 155, p. 62-70, 2017. <https://doi.org/10.1016/j.jastp.2017.02.002>.

WMO. Guide to Meteorological Instruments and Methods of Observation. **World Meteorological Organization**, 2014.

ZHOU, Y. LIU, Y. WANG, D. LIU, X. WANG, Y. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. **Energy Conversion and Management**, v. 235, n. 1, 2021. <https://doi.org/10.1016/j.enconman.2021.113960>.

3.0 CAPÍTULO 1 - ESTIMATIVAS DIÁRIAS DA RADIAÇÃO GLOBAL NA AMAZÔNIA BRASILEIRA POR MODELOS SIMPLIFICADOS

RESUMO: A radiação solar é um elemento e um fator meteorológico que está presente em diversos processos, como na evapotranspiração, fotossíntese, geração de energia e entre outros, porém para o seu conhecimento em algumas regiões, ocorre limitação de dados medidos de superfície. Neste trabalho foram avaliados 87 modelos empíricos para estimar a radiação global (Hg) no bioma Amazônico brasileiro; esses modelos foram divididos em cinco grupos de acordo com as variáveis de entrada, como a insolação (grupo I), temperatura do ar (grupo II), umidade relativa do ar (grupo III), variáveis astronômicas (grupo IV) e modelos híbridos (grupo V) como variável de entrada. Avaliou-se as estimativas primeiramente pela sua significância (Teste-t), e depois conforme as métricas estatísticas do desempenho dos modelos (R^2 , MBE, RMSE e d). O modelo do grupo V [$Hg/H_o = a + b \ln\Delta T + c S/S_o^d$] apresentou o melhor desempenho estatístico, em todas as indicativas avaliadas, seguido do modelo do grupo I [$Hg/H_o = a + b S/S_o^c$] e depois do modelo do grupo II [$Hg/H_o = a + b\Delta T + c\Delta T^{0,25} + d\Delta T^{0,5} + eT_{med}/H_o$]. Já os modelos do grupo III apresentaram baixo desempenho estatístico e os modelos grupo IV não foram significativos (NS) para todas as estações meteorológicas avaliadas. No geral, constatou-se que os modelos simples ou híbridos baseados na insolação e temperatura do ar se mostraram eficientes na estimativa da Hg no Bioma Amazônico brasileiro.

PALAVRAS-CHAVE: radiação solar; energia solar; dados meteorológicos mínimos; modelos empíricos de estimativas; indicador estatístico; bioma Amazônico.

3.1 INTRODUÇÃO

A radiação solar é uma fonte de energia limpa, abundante, contínua, renovável e com grande potencial de expansão da sua utilização no Brasil e no mundo. No contexto atual, a variabilidade espacial e temporal da incidência da radiação global (Hg) na superfície terrestre, pode ser considerada como um dos principais impulsionadores do seu uso, principalmente no contexto das mudanças climáticas (QIU et al. 2022; SAMANTA et al. 2022). Essas variações, segundo Souza et al. (2023), dependem de movimentos da Terra (rotação e translação), de fatores

geográficos (latitude, longitude, altitude, orientação e inclinação da superfície) e de fatores atmosféricos (sobretudo pela interação com gases de efeito estufa e vapor d'água) e nuvens.

De forma resumida, ao atravessar a atmosfera, as ondas eletromagnéticas que formam a radiação solar, podem sofrer atenuações pelos fenômenos da reflexão, absorção e difusão em função do contato com os gases constituintes da atmosfera, nuvens e/ou com partículas em suspensão; todavia, esses fenômenos não são suficientes para reter todo o espectro de ondas eletromagnéticas da radiação solar, e, um percentual atravessa a atmosfera sem qualquer interação (radiação direta), que, em conjunto com a componente difusa (seletiva e não seletiva, dependendo das dimensões da partícula atenuante) incidem a superfície do solo formando a radiação global (SOUZA et al. 2023).

O conhecimento da radiação global incidente em superfícies naturais e/ou forçadas com diferentes ângulos de inclinação e faces de exposição, pode ser aplicado em muitas áreas científicas e tecnológicas. Informações precisas sobre Hg são amplamente necessárias em diversas aplicações e processos químicos, físicos e biológicos (CHEN & LI, 2013) e em vários setores, como a energia renovável, meteorologia, agricultura, hidrologia, ecologia, conforto ambiental, epidemiologia, indústria, dentre outros (FAN et al. 2018; PIETRO & GARCIA, 2022).

Apesar da sua importância, em muitas regiões, principalmente nos países subdesenvolvidos, a medição da Hg ainda é um desafio devido aos altos custos dos instrumentos (sensores) e suas necessidades periódicas de recalibração e de manutenção; essas atividades são recorrentes e necessárias nas estações meteorológicas, mesmo que automáticas (HASSAN et al. 2016). Além disso, os conjuntos de dados de Hg são frequentemente caracterizados por uma quantidade considerável de falhas de medição ou com dados inconsistentes, decorrentes principalmente da falta de manutenção (sujeira depositada nos sensores, sombreamentos e falta de calibração dos sensores). Esses problemas ocorrem mesmo em países com elevada densidade de estações de monitoramento de Hg (CHEN & LI, 2013; ZHANG et al. 2017; YILDIRIM et al. 2018). Portanto, grandes esforços têm sido realizados para estimar Hg com base em variáveis meteorológicas que apresentam

menores custos de monitoramento e que são amplamente disponíveis e monitoradas nas estações meteorológicas em todo o mundo (QIU et al. 2022).

No Brasil, essa realidade não é diferente, especialmente, no bioma Amazônico, que se localiza ao Norte do país e ocupa 49% do território nacional, sendo que essa região apresenta a menor relação entre estações meteorológicas automáticas (EMAs) e estações meteorológicas convencionais (EMCs) por área da rede de estações do Instituto Nacional de Meteorologia (INMET). Além do número baixo de estações instaladas para monitoramento de superfície, os percentuais de falhas nas bases de dados existentes variam de 10 a 60% (SOUZA et al., 2017; BENDER & SENTELHAS, 2018; MARTIM & SOUZA, 2021), limitando assim, avaliações sazonais do comportamento de Hg incidente. E, mesmo em países que investem mais em pesquisas, esses problemas são observados, conforme relatado por Fan et al. (2018), avaliando 122 estações meteorológicas na China, observaram falhas de 0,3 a 7,5% nas bases de dados.

A caracterização espaço-temporal da radiação global requer uma série de medições de longo prazo, e, apesar da importância e do interesse da comunidade científica e governamental, o seu conhecimento ainda é incipiente no bioma Amazônia. Isso se deve, principalmente em função da grande extensão territorial, das limitações de acesso terrestre e da baixa densidade demográfica (número de cidades). Informações dos níveis diários de Hg nesse bioma são necessárias para diversas aplicações associadas às ciências ambientais, visto que evidências científicas sólidas indicam que uma parte representativa deste ecossistema está sendo afetada pelas ações antropogênicas, particularmente pela expansão das atividades agropecuárias, pela crescente frequência e gravidade dos incêndios florestais e pela degradação e redução da vegetação natural (DELGADO et al. 2022; SILVA JUNIOR et al. 2022; ARÉVALO et al. 2023).

A fim de atender às necessidades relevantes de obtenção de dados de radiação global, vários modelos foram desenvolvidos para estimativas de Hg com base em outros dados meteorológicos, e, vêm sendo comumente disponibilizadas/avaliadas em todo o mundo (JAHANI et al. 2017; CHEN et al. 2019; PRIETO & GARCIA, 2022). E não é diferente para regiões como a Amazônia brasileira, em que, apesar de alguns estudos já realizados (SOUZA et al., 2017; MARTIM & SOUZA, 2021), ainda carece

avaliar modelos simplificados de estimativas de Hg para construção de séries contínuas de dados para aplicações agropecuárias e ambientais.

As avaliações de modelos simplificados e/ou parametrizados de estimativas de Hg, demandam calibrações regionais de coeficientes estatísticos e que podem fornecer boas estimativas através de relações indiretas com outras variáveis meteorológicas, desde que estas sejam simples de se monitorar, apresentem maior disponibilidade de bases de dados ou boa correlação com Hg, como são os casos da insolação, temperatura do ar, umidade relativa do ar e da precipitação pluvial (ABRAHA & SAVAGE, 2008; CHEN & LI, 2013; HASSAN et al. 2016; FAN et al. 2018; PRIETO & GARCÍA, 2022).

O primeiro modelo empírico para estimativa da Hg baseado na insolação (S) foi proposto por Angström (1924), e, posteriormente modificado por Prescott (1940); em homenagem aos autores, ficou conhecido como modelo de Angstrom-Prescott (A-P). Já em relação a temperatura do ar, o primeiro modelo simplificado foi proposto por Hargreaves & Samani (1982), ao comprovarem que a amplitude térmica diária ($\Delta T = T_{\max} - T_{\min}$) apresenta boas correlações com Hg incidente; posteriormente Bristow & Campbell (1984) provaram que existe correlação exponencial entre a ΔT e Hg.

Trabalhos com base na insolação apresentam um desempenho superior quando comparado aos trabalhos baseados em outras variáveis meteorológicas, como por exemplo a temperatura do ar (ADARAMOLA, 2012; FALAYI et al. 2008); porém, no contexto brasileiro, a disponibilidade de dados de insolação é menor para a maioria dos biomas e/ou estados, ao contrário da temperatura do ar que é medida em praticamente todas as EMCs e EMAs (SANTOS et al. 2021; QIU et al. 2022); essa disparidade é mais acentuada no contexto da Amazônia brasileira. Por conseguinte, existem diversas possibilidades de ajustes de modelos baseados em outras variáveis meteorológicas ou geográficas que possuam correlação com Hg, pois isso permite aumentar as possibilidades de aplicações para diferentes situações de acordo com a disponibilidade dos dados, além de melhorar a capacidade preditiva da Hg.

Após a ampla difusão dos modelos de Angstrom-Prescott (1940), Hargreaves & Samani (1982) e Bristow & Campbell (1984), diversos autores propuseram modificações como ajustes polinomiais, logarítmicos, potenciais e exponenciais. Na revisão de literatura, Pietro & García (2022) relataram a existência de 165 modelos

distintos, que podem ser classificados de acordo com a variável meteorológica e não meteorológica utilizada: i) modelos baseados na insolação (S), em horas (Grupo I); ii) modelos baseados na temperatura máxima, média e mínima do ar, em °C (Grupo II); e iii) modelos que consideram a combinação de variáveis como insolação, temperatura máxima, média e mínima do ar, umidade relativa do ar, latitude, pressão atmosférica, declinação solar, velocidade do vento e precipitação pluvial (Grupo III).

Modelos empíricos vêm sendo estudados em diferentes regiões do mundo para estimativas de Hg na escala diária, sobretudo com base nesses três agrupamentos supracitados. Destaques podem ser dados à estudos baseados na insolação (S) para a China (CHEN & LI, 2013; CHEN et al. 2004; LI et al. 2013(B); NEWLAND 1989), Espanha (ALMOROX & HONTORIA, 2004), Turquia (BAKIRCI, 2009; ÖGELMAN et al. 1984; TOGRUL & ONAT, 1999; TOGRUL et al. 2000), Sudão (ELAGIB & MANSELL, 2000), Egito (EL-METWALLY, 2005), Arábia Saudita (EL-SEBAII et al. 2009), Nigéria (FALAYI et al. 2008), Irã (JAHANI et al. 2017), Coreia do Sul (LEE, 2015). Por conseguinte, outros estudos envolvendo modelos baseados na temperatura do ar (máxima, média e mínima diária) foram desenvolvidos para diferentes locais, por diferentes autores, como: Egito (HASSAN et al. 2016), China (CHEN & LI, 2013; FAN et al., 2018; CHEN et al., 2004; LI et al., 2013(A); LI et al., 2014; LI et al., 2015; QIU et al. 2022), Estados Unidos (ABRAHA & SAVAGE, 2008; BRISTOW & CAMPBELL, 1984; GOODIN et al. 1999; THORNTON & RUNNING, 1999; WEISS et al. 2001), Itália (ABRAHA & SAVAGE, 2008), Reino Unido (ABRAHA & SAVAGE, 2008), Holanda (ABRAHA & SAVAGE, 2008), África do Sul (ABRAHA & SAVAGE, 2008), Austrália (ABRAHA & SAVAGE, 2008), Nigéria (ADARAMOLA, 2012; FALAYI et al. 2008; KOLEBAJE et al. 2016), Arábia Saudita (BENGHANEM & MELLIT, 2014), África (HARGREAVES et al. 1985), Irã (JAHANI et al. 2017; SAFFARIPOUR et al. 2013), Índia (PANDAY & KATIYAR, 2010; SAMANTA et al. 2021), Turquia (YILDIRIM et al. 2018), modelos baseados na umidade relativa foram na Nigéria (ADARAMOLA, 2012), Barém (ELAGIB et al. 1998), Turquia (ERTEKIN & YALDIZ, 1999), Nigéria (FALAYI et al. 2008; KOLEBAJE et al. 2016) e modelos baseados em variáveis astronômicas na Jordânia (AL-SALAYMEH, 2006), Turquia (ERTEKIN & YALDIZ, 1999; TOGRUL & ONAT, 1999), Nigéria (FALAYI et al. 2008) e modelos híbridos para a China (CHEN

& LI, 2013, FAN et al. 2018; CHEN et al. 2004; LI et al. 2013(B); LI et al. 2015), Nigéria (ADARAMOLA, 2012; FALAYI et al. 2008; KOLEBAJE et al. 2016; ODODO et al. 1995; SWARTMAN & OGUNLADE, 1967), Sudão (ELAGIB & MANSELL, 2000), Barém (ELAGIB et al. 1998), Arábia Saudita (EL-SEBAII et al. 2009), Turquia (ERTEKIN & YALDIZ, 1999; YILDIRIM et al. 2018) e Coreia do Sul (LEE, 2015).

No bioma Amazônico, o estudo mais abrangente envolvendo estimativas de Hg foi realizado com o modelo de Angstrom-Prescott para 20 EMs (MARTIM & SOUZA, 2021). Trabalhos desenvolvidos em diversas regiões do Brasil, também apresentaram estimativas pontuais para algumas localidades da Amazônia (SOUZA et al., 2017; RAMOS et al., 2018; BENDER & SENTELHAS, 2018). Dada importância ambiental desse bioma e da Hg para diversas aplicações agroambientais, o objetivo deste estudo foi avaliar e determinar quais modelos baseados em variáveis meteorológicas e astronômicas apresentam melhor desempenho na estimativa da radiação global diária (Hg), na Amazônia brasileira, utilizando variáveis meteorológicas e astronômicas.

3.2 MATERIAL E MÉTODOS

3.2.1 Área de estudo

A área de estudo se concentra no bioma Amazônico brasileiro, que engloba uma extensa área de 4.196.943 km², distribuída em 9 estados: Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e parte do Tocantins, Mato Grosso e Maranhão (Figura 3.1). Foram selecionados 20 locais com estações meteorológicas (EMs) que apresentavam medidas concomitantes de forma automáticas (EMAs) e convencionais (EMCs) (Tabela 3.1), sendo consideradas as seguintes bases de dados diárias: i) das EMAs - temperatura máxima, média e mínima (Tmax, Tmed e Tmin), umidade relativa média (RHmed) e radiação global horizontal (Hg); ii) das EMCs - insolação (S). Essas bases de dados foram obtidas no site do Instituto Nacional de Meteorologia (INMET) (<https://portal.inmet.gov.br/>) (INMET, 2022). Embora a rede de estações do INMET seja maior do que a selecionada nesse estudo para a região Amazônica, destaca-se que foram selecionadas apenas àquelas que apresentavam medidas concomitantes da Hg e S.

Segundo a Classificação de Köppen, as 20 estações estão inseridas em três sub-classes de climas tropicais: Am (monção), Af (com estação seca) e Aw (com inverno seco), com alta precipitação (superiores a 3600 mm ano^{-1}) ao Norte do bioma e com redução em direção ao Sul da Amazônia (ALVARES et al. 2013). Geograficamente, as EMAs e EMCs estão inseridas entre as latitudes $0,035^\circ$ a $-11,97^\circ$ e longitudes $-48,43^\circ$ a $-69,87^\circ$ (três fusos horários distintos), enquanto as altitudes variam de 9 a 366 metros. A série de dados mais longa avaliada foi de 22 anos (2000 a 2022) na estação de Manaus (AM) e a série mais curta de dados foi de 5 anos (2012 a 2017) na estação de Soure (PA) (Tabela 3.1).

Os dados meteorológicos foram submetidos à filtros para identificação de inconsistências e falhas, sendo que nesse caso, foram excluídos os dias que apresentaram falhas diárias de Tmax, Tmed, Tmin, S, RHmed e falhas horárias da Hg entre as 10 e 14 horas (horário solar local); também foram excluídos os dias com valores de razão de insolação (R_i) $> 1,0$ e $K_t > 0,85$. Posteriormente, por EMA, a série de dados foi dividida em 70% para calibração dos coeficientes dos modelos e o restante (30%) foi empregada para avaliação do desempenho estatístico; essa separação das bases de dados foi realizada garantindo a representatividade e proporcionalidade dos períodos do ano (meses), em ambas as bases de dados.

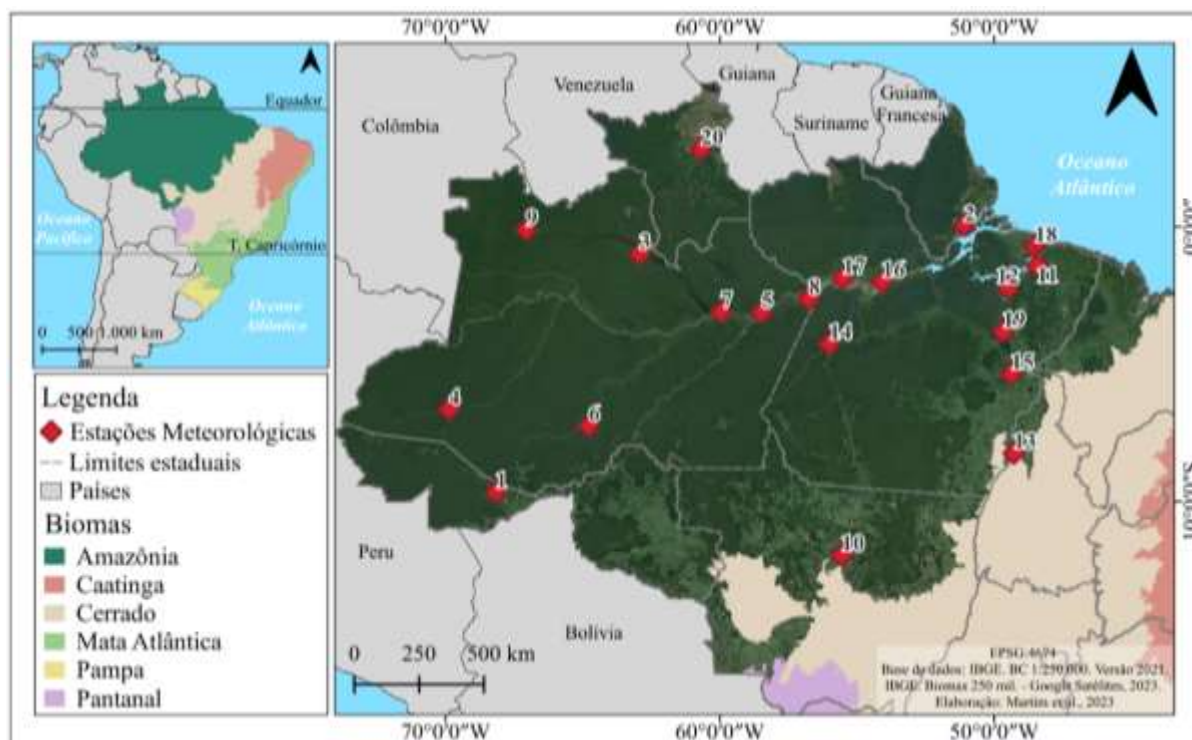


Figura 3.1. Localização das 20 estações meteorológicas avaliadas neste estudo, no bioma Amazônia, Brasil. (A ordem numérica das estações está apresentada na Tabela 1).

Tabela 3.1. Estações meteorológicas na Amazônia brasileira e suas respectivas localizações, classificação climática, coordenadas geográficas e período de dados.

Estado	Cidade	CCKP*	Lat.	Lon.	Alt.	Período de funcionamento
Acre (AC)	1 - Rio Branco	Am	-9,67	-68,16	163	2015-2022
Amapá (AP)	2 – Macapá	Am	0,035	-51,08	16	2013-2022
Amazonas (AM)	3 – Barcelos	Af	-0,98	-62,92	29	2008-2022
	4 – Eirunepé	Af	-6,65	-69,87	121	2012-2022
	5 – Itacoatiara	Af	-3,12	-58,47	41	2008-2022
	6 – Lábrea	Am	-7,25	-64,78	61	2008-2018
	7 – Manaus	Af	-3,1	-59,95	61	2000-2022
	8 – Parintins	Af	-2,63	-56,75	18	2008-2018
	9 - São Gabriel da Cachoeira	Af	-0,12	-67,05	79	2011-2022
Mato Grosso (MT)	10 – Sinop	Aw	-11,85	-55,55	366	2006-2017
Pará (PA)	11 – Belém	Af	-1,41	-48,43	21	2003-2022
	12 – Cametá	Af	-2,23	-49,48	9	2008-2022
	13 - Conceição do Araguaia	Aw	-8,25	-49,27	175	2008-2022
	14 – Itaituba	Af	-4,27	-56,00	24	2008-2022
	15 – Marabá	Aw	-5,36	-49,37	116	2009-2022
	16 - Monte Alegre	Am	-2	-54,07	100	2012-2022
	17 - Óbidos	Am	-1,88	-55,51	89	2012-2017
	18 - Soure	Am	-0,72	-48,51	12	2008-2017
19 - Tucuruí	Am	-3,82	-49,67	137	2008-2017	
Roraima (RR)	20 - Boa Vista	Am	2,82	-60,68	82	2010-2022

Latitude (Lat.); Longitude (Lon.); Altitude (Alt.); *Fonte: Alvares et al. (2013)

3.2.2 Modelos simplificados da estimativa de HG

Todos os modelos simplificados avaliados, baseiam-se na radiação extraterrestre (H_0), que corresponde a radiação que ainda não interagiu com os elementos atmosféricos; essa variável foi obtida pela Equação 3.1, que depende do fator de excentricidade da órbita da Terra (dr - equação 3.2), declinação solar (δ - equação 3.3) e do ângulo horário diário (h - equação 3.4), sendo que, além de H_0 , alguns modelos simplificados dependem do fotoperíodo (S_0 - equação 3.5) ou da transmissividade atmosférica (K_t - que representa a fração de radiação que chega na superfície terrestre - equação 3.6) ou da razão de insolação (R_i - que representa a fração de tempo de

incidência de radiação direta ao longo de um dia - equação 07) (ALLEN et al. 1998; SOUZA et al. 2017).

$$H_o = 37,59 * dr * \left(\frac{\pi}{180} * h * \sin \phi * \sin \delta + \cos \phi * \cos \delta * \sin h \right) \quad 3.1$$

$$dr = 1 + 0,033 * \cos \left(\frac{360 * DJ}{365} \right) \quad 3.2$$

$$\delta = 23,45 * \sin \left[\frac{360}{365} (DJ + 284) \right] \quad 3.3$$

$$h = \cos^{-1}(-\tan \phi * \tan \delta) \quad 3.4$$

$$S_o = \frac{2 * \sigma}{15} \quad 3.5$$

$$K_t = \frac{H_g}{H_o} \quad 3.6$$

$$R_i = \frac{S}{S_o} \quad 3.7$$

em que: ϕ é a latitude local que as estações estão localizadas($^{\circ}$); δ e dr são dependentes apenas da época do ano (DJ, que indica o ordenamento dos dias ao longo do ano ($1 \leq DJ \leq 365$ ou 366 dias)); h é o ângulo horário que depende da época do ano e da latitude; os valores de K_t e R_i são adimensionais.

Com base na revisão de literatura, foram obtidos 87 modelos simplificados para estimativas diárias de H_g , que apresentam diferenças nas variáveis de entrada ou no modelo analítico de estimativa. Esses modelos foram divididos em 5 Grupos de acordo com as variáveis de entrada, gerando as seguintes combinações:

- - Grupo I: modelos empíricos baseados na insolação (S), fotoperíodo (S_o) e radiação extraterrestre (H_o) – Tabela 3.2;
- - Grupo II: modelos empíricos baseados na temperatura máxima, média e mínima do ar (T_{max} , T_{med} , T_{min}) e radiação extraterrestre (H_o) – (Tabela 3.3);
- - Grupo III: modelos empíricos baseados na umidade relativa do ar (RH) e radiação extraterrestre (H_o) – (Tabela 3.4);
- - Grupo IV: modelos empíricos com variáveis astronômicas, como a declinação solar (δ), dia juliano (DJ) e radiação extraterrestre (H_o) – (Tabela 3.4);
- - Grupo V: modelos empíricos híbridos, baseados em duas ou mais variáveis como S, T_{max} , T_{med} , T_{min} , RHmed e latitude (ϕ), fotoperíodo (S_o) e radiação extraterrestre (H_o) (Tabela 3.5).

As calibrações locais das estimativas dos coeficientes empíricos (“a”, “b”, “c” e “d”) dos modelos simplificados foram realizadas para cada estação com o agrupamento de dados, utilizando o algoritmo iterativo de Gauss-Newton para

otimizações não lineares pelo método dos mínimos quadrados, no software Statistica 14.0.0.15 (TIBCO Software Inc), juntamente com avaliação da maximização dos coeficientes de determinação (R^2) e restrições para $H_g < H_o$. Sequencialmente, são apresentadas análises para os agrupamentos dos modelos.

3.2.3 Modelos baseados na insolação – GRUPO I

O modelo de Angstrom-Prescott foi o precursor desse tipo de estimativa (Tabela 3.2), e correlaciona linearmente a razão de insolação (R_i) e a transmissividade atmosférica (K_t); nesse caso, o coeficiente linear (“a”) e angular (“b”) são dependentes das condições atmosféricas regionais e precisam ser calibrados localmente. Embora estes coeficientes sejam empíricos, eles descrevem duas possibilidades físicas de incidência de H_g : i) quando R_i tende a zero, o coeficiente linear (a) indica a transmissividade mínima da atmosfera da região, dada pela radiação difusa; ii) quando R_i tende 1,0, a soma dos dois coeficientes indica a transmissividade máxima da atmosfera da região em dias de céu claro, e as contribuições da radiação difusa e direta (ALMOROX & HONTORIA, 2004).

Essa metodologia está baseada na obtenção de S por meio de heliógrafos de Campbell-Stockes, posicionados e alinhados em função da latitude local, e que apresentam concentração e projeção de raios solares nos heliogramas, em função da movimentação diurna aparente do Sol (ângulo zenital). Apesar de simplificada, essa análise pode indicar alterações no padrão de incidência da radiação direta em função do aumento da multireflexão atmosférica e do brilho horizontal quando S tende a 0,0 (AL-SALAYMEH, 2006). Além disso, ressalta-se que pode ocorrer incidência de radiação global em nível abaixo do necessário para a queima dos heliogramas (estimada em torno 120 Wm^{-2}), todavia, esses níveis energéticos podem ser computados como radiação difusa em condições de céu nublado (WMO, 2014).

Os modelos 2 a 13 foram desenvolvidos como alterações do modelo de Angstrom-Prescott, com a inserção de termos polinomiais de segundo e terceiro grau, logaritmos, exponenciais e potenciais, e, são funções das variáveis S , S_o e H_o . Houve a necessidade da inserção da expressão $f(S) = (S + 1)$ no lugar do S representado nos modelos 4, 6 e 11 pois são modelos limitados em dias onde a insolação é nula ($S = 0$), sendo que essa condição representa 4,5% da série histórica das estações meteorológicas localizadas no bioma Amazônico.

Tabela 3.2. Modelos empíricos de estimativas indiretas de radiação global baseados na insolação.

Nº	Referências	Modelo
1	Angström (1924); Prescott (1940)	$Hg/Ho = a + b (S / So)$
2	Ögelman et al. (1984)	$Hg/Ho = a + b (S / So) + c (S / So)^2$
3	Bahel (1987)	$Hg/Ho = a + b (S / So) + c (S / So)^2 + d (S / So)^3$
4	Newland (1989)	$Hg/Ho = a + b (S / So) + c \ln (f(S) / So)$
5	Togrul & Onat (1999)	$Hg/Ho = a + b/Ho + c (S / So) / Ho$
6	Togrul et al. (2000)	$Hg/Ho = a + b \ln (f(S) / So)$
7	Almorox & Hontoria (2004)	$Hg/Ho = a + b \text{EXP} (S / So)$
8		$Hg/Ho = a \text{EXP} (b S / So)$
9	Elagib & Mansell (2000)	$Hg/Ho = a (S / So)^b$
10		$Hg/Ho = a + b (S / So)^c$
11	El-Metwally (2005)	$Hg/Ho = a^{1/(f(S) / So)}$
12	Bakirci (2009)	$Hg/Ho = a + b \text{EXP} (S / So) + c (S / So)$
13	Li et al. (2013) B	$Hg/Ho = a + b/Ho + c S / Ho$

3.2.4 Modelos baseados na temperatura do ar – GRUPO II

Geralmente as medições de insolação (S) estão limitadas a poucas estações meteorológicas (CHEN & LI, 2013; ABRAHA & SAVAGE, 2008); como por exemplo, na China, a “S” é monitorada em apenas 30% das estações do país (QIU et al. 2022); especificamente no bioma Amazônico (região deste estudo), têm-se em torno de 27%. Nos países em desenvolvimento, normalmente, a temperatura do ar é a variável meteorológica com maior difusão espacial e temporal de medidas, em função da simplicidade dos sensores e por sua ampla aplicação (BENGHANEM & MELLIT, 2014).

Os modelos 14 a 45 (Tabela 3.3) são baseados em diferentes correlações diárias entre a Hg e variações estatísticas da temperatura do ar como, temperatura máxima (Tmax), temperatura média (Tmed), temperatura mínima (Tmin), amplitude térmica média mensal (ΔT_{med}), amplitude térmica diária ($\Delta T = Tmax - Tmin$); sendo que esses modelos ainda podem ser classificados em 5 sub-grupos, em função das diferentes combinações de variáveis de entrada: (Tmax, Tmin, Ho), (Tmed, Ho), (Tmin, Ho), (Tmax, Ho) e (Tmax, Tmed, Tmin, Ho).

Tabela 3.3. Modelos empíricos de estimativas diárias da radiação global baseados na temperatura do ar.

Nº	Referências	Modelo
14	Hargreaves & Samani (1982)	$Hg/Ho = a \Delta T^{0.5}$
15	Bristow & Campbell (1984)	$Hg/Ho = a (1 - EXP(-b \Delta T^c))$
16	Hargreaves et al. (1985)	$Hg/Ho = a + b \Delta T^{0.5}$
17	Ertekin & Yaldiz (1999)	$Hg/Ho = a + b / Ho + c Tmed / Ho$
18	Goodin et al. (1999)	$Hg/Ho = a (1 - EXP(-b \Delta T^c / Ho))$
19	Thornton & Running (1999)	$Hg/Ho = 1 - EXP(- a \Delta T^b)$
20	Weiss et al. (2001)	$Hg/Ho = 0.75 (1 - EXP(- a \Delta T^2 / Ho))$
21	Chen et al. (2004)	$Hg/Ho = a + b \ln \Delta T$
22	Abraha & Savage (2008)	$Hg/Ho = 0.75 (1 - EXP(- a \Delta T^2 / \Delta Tmed))$
23		$Hg/Ho = a + b Tmin$
24	Falayi et al. (2008)	$Hg/Ho = a + b Tmax$
25		$Hg/Ho = a + b Tmax / Tmin$
26	Panday & Katiyar (2010)	$Hg/Ho = a + b Tmax / Tmin + c (Tmax / Tmin)^2$
27		$Hg/Ho = a + b Tmax / Tmin + c (Tmax / Tmin)^2 + d (Tmax / Tmin)^3$
28	Adaramola (2012)	$Hg/Ho = a + b Tmed$
29		$Hg/Ho = a + b Tmin / Tmax$
30		$Hg/Ho = a + b \Delta T$
31	Chen & Li (2013)	$Hg/Ho = a + b Tmin + c Tmax + d Tmin Tmax$
32		$Hg/Ho = a + b Tmin + c Tmax$
33	Li et al. (2013)	$Hg/Ho = a/Ho + b Tmin + c Tmax$
34	Benghanem & Mellit (2014)	$Hg/Ho = a/Ho + b \Delta T^c$
35	Li et al. (2014)	$Hg/Ho = a + (b + c Tmed) \Delta T^{0.5}$
36		$Hg/Ho = a + b \Delta T^c$
37		$Hg/Ho = a + b Ho Tmed^c$
38	Hassan et al. (2016)	$Hg/Ho = a Ho Tmed^b$
39		$Hg/Ho = a EXP(b Tmed^c)$
40		$Hg/Ho = a + b Tmed + c Tmed^2$
41		$Hg/Ho = (a + b \Delta T + c \Delta T^2) \Delta T^d$
42	Jahani et al. (2017)	$Hg/Ho = a + b \Delta T + c \Delta T^2 + d \Delta T^3$
43		$Hg/Ho = a + b \Delta T^{0.5} + c \Delta T^{1.5} + d \Delta T^{2.5}$
44	Fan et al. (2018)	$Hg/Ho = a + b \Delta T + c \Delta T^{0.25} + d \Delta T^{0.5}$
45		$Hg/Ho = a + b \Delta T + c \Delta T^{0.25} + d \Delta T^{0.5} + e Tmed / Ho$

3.2.5 Modelos baseados na umidade relativa do ar ou em variáveis astronômicas – GRUPO III e IV

Na literatura é observado menor número de modelos desenvolvidos para essas variáveis de entrada, visto que as suas relações com a radiação global são indiretas e com maiores intervalos temporais de resposta com a Hg. A umidade relativa do ar pode ser obtida por conjuntos psicrométricos (bulbo seco e bulbo úmido) ou por sensores capacitativos, e, nesse caso, nas condições brasileiras, ocorre maior disponibilidade de valores médios diários, sendo os modelos 46 a 49 representam esse grupo. Por sua vez, os modelos 50 a 54 são baseados em variáveis astronômicas, que geralmente estão diretamente associados à época do ano, em função da declinação solar (δ) e do ordenamento numérico dos dias ao longo do ano (DJ – dia juliano) e radiação extraterrestre (H_o) (Tabela 3.4).

Tabela 3.4. Modelos empíricos de estimativas diária da radiação global baseados na umidade relativa do ar (RHmed) e com variáveis astronômicas (δ , H_o e DJ).

Nº	Referências	Modelo
46	Elagib et al. (1998)	$Hg/H_o = a/H_o + b \text{ RHmed}/H_o$
47		$Hg/H_o = a/H_o + b (\text{RHmed} - H_o)/H_o$
48	Falayi et al. (2008)	$Hg/H_o = a + b \text{ RHmed}$
49	Kolebaje et al. (2016)	$Hg/H_o = a + b \text{ RHmed}^{0.5}$
50	Ertekin & Yaldiz (1999)	$Hg / H_o = a / H_o + b \delta / H_o$
51	Togrul & Onat (1999)	$Hg / H_o = a / H_o + b \text{ sen } \delta / H_o$
52	Togrul & Onat (1999)	$Hg / H_o = a + b / H_o$
53	Al-Salaymeh (2006)	$Hg / H_o = a / H_o + b \text{ sen } (2 \pi \text{ DJ} / c + d) / H_o$
54	Al-Salaymeh (2006)	$Hg / H_o = (a + b \text{ DJ} + c \text{ DJ}^2 + d \text{ DJ}^3 + e \text{ DJ}^4) / H_o$

3.2.6 Modelos empíricos híbridos – GRUPO V

Os modelos complexos ou híbridos objetivam melhorar o desempenho dos modelos empíricos simplificados (JAHANI et al. 2017), pois além da influência direta da radiação global na temperatura do ar ou na insolação, existem outros fatores astronômicos e geográficos que podem interferir diretamente na incidência de radiação na superfície terrestre (BENGHANEM & MELLIT, 2014). Na Tabela 3.5 são apresentados 33 modelos dependentes das relações entre ϕ , H_o , S, S_o , T_{max} , T_{med} , T_{min} , RHmed , ΔT e δ .

Tabela 3.5. Modelos empíricos híbridos de estimativas diárias da radiação global com variáveis meteorológicas e astronômicas.

Nº	Referências	Modelo
55	Glover & McCulloch (1958)	$Hg/Ho = a \cos \phi + b (S / So)$
56		$Hg/Ho = a \text{ EXP } (b (S / So - RHmed))$
57	Swartman & Ogunlade (1967)	$Hg/Ho = a + b RHmed + c S / So$
58		$Hg/Ho = a RHmed^b (S / So)^c$
59	Ododo et al. (1995)	$Hg/Ho = a Tmed^b RHmed^c (S / So)^d$
60		$Hg/Ho = a + b Tmed + c RHmed + (d Tmed + e) S / So$
61	Elagib et al. (1998)	$Hg/Ho = a/Ho + b (RHmed - \Delta T - Ho)/Ho$
62	Chen et al. (2004)	$Hg/Ho = a + b \ln \Delta T + c (S / So)^d$
63		$Hg/Ho = a + b Tmed + c (S / So)$
64	Falayi et al. (2008)	$Hg/Ho = a + b Tmin + c (S / So)$
65		$Hg/Ho = a + b Tmax + c (S / So)$
66		$Hg/Ho = a + b Tmed + c RHmed + d (S / So)$
67	El-Sebaï et al. (2009)	$Hg/Ho = a + b Tmed + c RHmed$
68	Adaramola (2012)	$Hg/Ho = a + b (Tmin/Tmax) RHmed / 100$
69	Korachagaon & Bapat (2012)	$Hg/Ho = a + b Tmax + c \Delta T + d RHmed$
70		$Hg/Ho = a + b Tmax + c Tmin + d \Delta T + e RHmed$
71		$Hg/Ho = a + b \Delta T^{0.5} + c (S / So)$
72	Chen & Li (2013)	$Hg/Ho = a + b Tmin + c Tmax + d (S / So)$
73		$Hg/Ho = a + b Tmax + c Tmin + d RHmed + e (S / So)$
74		$Hg/Ho = a + b \Delta T^{0.5} + c RHmed$
75		$Hg/Ho = a / Ho + b Tmin + c Tmax + d RHmed$
76	Li et al. (2013) B	$Hg/Ho = a / Ho + b Tmin + c Tmax + d RHmed / Ho$
77		$Hg/Ho = a / Ho + b \sqrt{(\Delta T)} + c RHmed$
78		$Hg/Ho = a / Ho + b \sqrt{(\Delta T)} + c (RHmed / Ho)$
79		$Hg/Ho = a + b / Ho + c Tmax / Ho + d (S / So) / Ho$
80	Saffaripour et al. (2013)	$Hg/Ho = a + b / Ho + c RHmed / Ho + d (S / So) / Ho$
81		$Hg/Ho = a + b \text{ sen } \delta + c (S / So)$
82	Lee (2015)	$Hg/Ho = a + b (S / So)^c + d \Delta T^e$
83	Li et al. (2015)	$Hg/Ho = a (1 + b RHmed) \Delta T^{0.5}$
84		$Hg/Ho = a (1 + b RHmed) (1 - \text{EXP}(-c \Delta T^d))$
85	Kolebaje et al. (2016)	$Hg/Ho = a + b \Delta T / f(S)$
86		$Hg/Ho = a + b ((\Delta T + RHmed) / So)^{0.5}$
87	Yildirim et al. (2018)	$Hg/Ho = a + b RHmed + c S / So + d (S / So)^2 + e (S / So)^3$

3.2.7 Desempenho estatístico dos modelos

Após as calibrações locais dos coeficientes de cada modelo, foi realizado o Teste-t (que é um teste de hipótese) para avaliar a significância (valor adotado de $p \leq 0,05$) do coeficiente do modelo (β_a) em relação ao erro padrão do coeficiente estimado

($SE_{\beta a}$), de forma a verificar a possibilidade de emprego e validação dos coeficientes nas estimativas dos modelos de Hg (equação 3.8).

$$\text{Teste } - t = \frac{\beta a}{SE_{\beta a}} \quad 3.8$$

Na literatura, existem vários indicativos estatísticos para avaliação do desempenho de modelos estatísticos paramétricos empregados em estimativas meteorológicas e climáticas, incluindo a obtenção de Hg (BADESCU, 2013, TEKE et al. 2015). Em geral, vêm sendo recomendados a utilização da raiz quadrada do erro quadrático médio (RMSE), erro relativo médio (MBE), índice “d” de Willmott e o coeficiente de determinação ou R^2 (equações 3.9 a 3.12 respectivamente).

Segundo Adaramola (2012), a análise MBE ocorre da seguinte maneira, quando positivo representa superestimativas, e, quando negativo indica subestimativas da variável estudada; sendo que a desvantagem desse indicativo é que valores altos com sinais diferentes podem acabar se anulando, gerando um erro pequeno, que prejudica a análise do modelo de estimativa em estudo. Já o índice RMSE, é representado apenas por valores positivos, e tem como referência valores próximos a zero como ideal; por ser função quadrática, grandes variações nos dados podem aumentar os valores, ou seja, é uma métrica de como está o espalhamento dos valores estimados em relação aos medidos. O índice de Willmott ou “d” é um indicativo que fornece valores adimensionais, entre 0 a 1, sendo que quanto maior o valor, melhor é o desempenho do modelo. O coeficiente de determinação ou R^2 indica a qualidade da relação ajuste linear entre os valores medidos (variável independente) e estimados (variável dependente), que varia de 0 a 1, sendo que valores mais próximos de 1 indicam melhores estimativas.

Embora existam vários indicativos estatísticos, não é recomendada de forma isolada a utilização de apenas um, índice de qualidade do ajuste, ou seja, devem ser utilizados diversos indicadores em conjunto, pois eles podem apresentar respostas distintas ou próximas quanto aos desempenhos estatísticos dos modelos analisados (BADESCU, 2013; FAN et al. 2018).

$$RMSE = \left[\frac{\sum_{i=1}^n (P_i - O_i)^2}{n} \right]^{\frac{1}{2}} \quad 3.9$$

$$MBE = \left[\frac{\sum_{i=1}^n (P_i - O_i)}{n} \right] \quad 3.10$$

$$d = 1 - \left[\frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - O_i| + |O_i - \bar{O}|)^2} \right] \quad 3.11$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad 3.12$$

em que: P_i é o valor estimado; O_i é o valor de referência observado nas estações meteorológicas; \bar{O} é a média dos valores de referência e n é o número de observações.

3.3 RESULTADOS

Embora no bioma Amazônico brasileiro predomine o clima tropical, foram observadas variabilidades entre as médias das variáveis meteorológicas nas diferentes estações avaliadas (Tabela 3.6). As médias anuais e os erros padrões das médias de Hg variaram de $15,1 \pm 3,59$ (Belém) até $20,6 \pm 4,19$ MJ m⁻² d⁻¹ (Monte Alegre), e geraram coeficientes de transmissividade atmosférica médios de $0,42 \pm 0,12$ (Belém) a $0,57 \pm 0,11$ (Monte Alegre); as médias anuais de insolação (S) variaram de $3,94 \pm 2,70$ (Eirunepé) a $7,57 \pm 2,59$ horas (Cametá), perfazendo em razões de insolação de $0,33 \pm 0,22$ (Eirunepé) a $0,63 \pm 0,21$ (Monte Alegre). Quanto às temperaturas do ar, foram observadas variações de $30,94 \pm 0,95$ (Soure) a $33,54 \pm 2,75$ °C (Conceição do Araguaia) para Tmax; de $25,41 \pm 1,63$ (Sinop) a $27,83 \pm 1,56$ °C (Roraima) para Tmed; e de $20,16 \pm 2,11$ (Sinop) a $25,34 \pm 1,51$ °C (Soure) para Tmin; por conseguinte, a RHmed anual variou de $68,54 \pm 10,17$ (Roraima) a $83,88 \pm 6,11\%$ (Barcelos), e, as precipitações pluviais totais anuais oscilaram de $1,616 \pm 100$ (Roraima) a 3.205 ± 129 mm (Belém). As médias anuais das variáveis teóricas como radiação no topo da atmosfera (Ho), apresentaram pequenas oscilações ($35,76 \pm 2,95$ a $36,36 \pm 2,55$ MJ m⁻² d⁻¹), em função das latitudes locais.

Ao se analisar as correlações entre as variáveis meteorológicas (Figura 3.2), de forma conjunta para as 20 estações meteorológicas, observa-se que para a radiação global, apenas na correlação com a umidade relativa média (RHmed) ocorre diminuição de Hg com aumento de RHmed; nesse caso, ao ajustar a regressão linear simples ($Hg = 37,74 - 0,2627 * RHmed$), obteve-se um coeficiente de correlação de -0,52. Por outro lado, as piores correlações são observadas com a Tmin ($r = 0,081$). Correlações fracas entre Hg e Tmin ou RHmax são esperadas, visto que existe uma interdependência dessas duas variáveis quanto aos instantes de ocorrência (normalmente no período noturno – ausência de radiação solar). Observa-se ainda

correlações mais fortes e crescentes de Hg com Tmed e Tmax com coeficientes de correlação de 0,56 e 0,66, respectivamente, já para a insolação (S) mensurada nos heliógrafos, verifica-se que existe um coeficiente linear que indica a existência de S para valores de Hg mínimos de aproximadamente $2,0 \text{ MJ m}^{-2} \text{ d}^{-1}$, com coeficiente de correlação de 0,83 e menores dispersões.

Quanto as distribuições das frequências, observa-se distribuições normais para Tmax, Tmed e Tmin; já para Hg e RHmed, as médias ocorrem em percentis superiores a 60%; e no tocante a S, têm-se maiores ocorrências percentuais de dias nublados ($S = 0$) e comportamento crescente da frequência de ocorrência entre 1 e 11 horas de insolação. As interpolações de Hg apresentam maior uniformidade quando associadas a Tmax e S.

Visando avaliar ao longo do ano o comportamento da radiação global em função da sazonalidade da precipitação pluvial no bioma Amazônia ao longo do ano, foi considerado o agrupamento de dados mensais para as 20 estações meteorológicas, sendo que os valores foram plotados em gráficos de boxplots (Figura 3.3).

Tabela 3.6. Médias anuais e desvio padrão das variáveis meteorológicas (Hg, S, Tmax, Tmed, Tmin, RHmed e Prec. – precipitação pluvial dos totais anuais médios) e astronômica (Ho) para as 20 estações meteorológicas estudadas na Amazônia brasileira (para precipitação pluvial considera-se os totais anuais médios).

Estações	Hg (MJ m ⁻² d ⁻¹)	Ho (MJ m ⁻² d ⁻¹)	Kt	S (horas)	Ri	Tmax (°C)	Tmed (°C)	Tmin (°C)	RHmed (%)	Prec. (mm ano ⁻¹)
1	17,17±4,80	36,23±3,40	0,47±0,13	5,58±3,11	0,46±0,26	31,29±2,84	25,60±2,04	21,68±1,96	78,42±12,52	2.954±139
2	19,86±5,28	36,12±1,35	0,55±0,14	6,95±3,23	0,58±0,26	31,76±1,65	27,54±1,22	23,97±0,73	76,56±1,22	2.100±145
3	17,17±5,23	35,99±1,34	0,48±0,14	4,77±3,12	0,40±0,08	32,02±2,29	26,34±1,23	22,76±1,18	83,88±6,11	2.443±72
4	15,64±4,25	36,36±2,55	0,43±0,12	3,94±2,70	0,33±0,22	31,55±2,27	25,92±1,48	22,24±1,39	70,16±14,29	1.952±75
5	16,12±5,09	36,05±1,98	0,45±0,14	5,78±3,33	0,48±0,27	31,52±2,24	27,24±1,44	24,01±0,98	79,57±6,57	2.339±104
6	17,15±3,84	35,76±2,95	0,48±0,10	5,24±3,30	0,44±0,25	32,75±2,10	26,70±1,30	22,57±1,51	78,86±5,96	2.230±103
7	16,34±5,04	35,91±2,03	0,46±0,14	5,52±3,23	0,46±0,27	32,30±2,21	27,74±1,64	24,32±1,22	75,86±9,16	2.206±99
8	17,52±5,41	35,88±1,84	0,49±0,14	6,17±3,41	0,51±0,28	31,29±2,07	27,15±1,43	24,24±1,09	81,09±6,72	2.343±110
9	15,22±4,76	36,17±1,30	0,42±0,12	4,73±2,81	0,39±0,23	31,30±2,23	26,41±1,45	23,14±1,19	81,46±7,99	2.867±46
10	19,13±4,19	35,95±3,96	0,53±0,12	6,03±3,04	0,50±0,26	32,35±2,81	25,41±1,63	20,16±2,11	72,04±15,78	1.952±132
11	15,09±3,59	36,04±1,55	0,42±0,10	6,48±2,75	0,54±0,23	32,67±1,35	27,27±1,09	23,56±0,65	78,49±5,75	3.205±129
12	20,16±3,78	35,91±1,79	0,56±0,10	7,57±2,59	0,63±0,21	32,47±1,21	27,75±1,13	24,23±1,02	74,36±6,15	2.230±137
13	18,64±4,46	35,79±3,26	0,52±0,13	6,96±3,26	0,58±0,28	33,54±2,75	26,83±1,69	21,60±2,12	70,50±12,26	1.686±104
14	18,75±4,71	36,03±2,25	0,52±0,13	6,24±3,18	0,52±0,26	32,67±2,17	27,58±1,46	23,85±0,96	74,87±7,16	2.069±95
15	18,25±3,87	35,82±2,57	0,51±0,11	6,36±3,10	0,53±0,26	32,26±1,95	26,59±1,14	22,40±1,37	76,53±7,75	1.885±123
16	20,61±4,19	36,13±1,71	0,57±0,11	7,53±2,79	0,63±0,23	31,66±1,69	27,54±1,29	23,97±1,05	75,30±6,98	1.661±104
17	16,64±4,52	36,21±2,31	0,46±0,12	6,70±3,21	0,56±0,26	33,08±2,45	26,84±1,46	22,74±0,78	78,22±8,71	2.572±107
18	19,82±4,30	35,96±1,38	0,55±0,12	6,89±3,55	0,57±0,29	30,94±0,95	27,71±1,04	25,34±1,51	76,98±6,03	2.093±74
19	16,95±3,48	36,06±1,99	0,47±0,09	6,22±2,81	0,52±0,23	31,43±1,68	26,73±1,15	23,36±0,94	78,42±7,70	2.400±157
20	19,35±4,35	35,99±1,77	0,54±0,11	6,49±2,87	0,54±0,23	33,51±2,22	27,83±1,56	23,70±1,07	68,54±10,17	1.616±100

Radiação global (Hg, MJ m⁻² d⁻¹), radiação extraterrestre (Ho, MJ m⁻² d⁻¹), coeficiente de transmissividade (Hg/Ho), insolação (S, horas), fotoperíodo (So, horas), razão de insolação (S/So), temperatura máxima (Tmax, °C), temperatura média (Tmed, °C), temperatura mínima (Tmin, °C), umidade relativa média (RHmed, %) e precipitação pluvial (Prec., mm ano⁻¹).

Embora muitas estações meteorológicas estejam inseridas em latitudes próximas, a radiação global média incidente pode variar significativamente entre si; como por exemplo nas estações meteorológicas de Macapá (0,035° N), Barcelos (-0,98° S) e São Gabriel da Cachoeira (-0,12° S), em que as medianas de Hg foram de 21,52; 18,09 e 15,79 MJ m⁻² d⁻¹, respectivamente; esse comportamento que mostra a influência das precipitações (nebulosidade) na Hg, visto que nessas estações meteorológicas, os totais anuais de chuvas são de 2100, 2433 e 2867 mm, respectivamente. Ainda pode-se considerar na comparação entre essas mesmas três estações meteorológicas, pode-se considerar que o efeito da proximidade de grandes superfícies de água livre (como ocorre em Macapá), a incidência de Hg depende dos movimentos de circulação atmosférica regional ou de mesoescala; em superfícies livres de água, potencialmente pode ocorrer um aumento das pressões atuais de vapor d'água (ea) na atmosfera, em função dos maiores níveis de evaporação direta da água, todavia, esse vapor d'água pode ser transportado na atmosfera para outras regiões em funções dos ventos, diminuindo assim a atenuação do vapor d'água na Hg em uma escala local.

Na estação chuvosa (outubro a abril), em geral as médias de Hg são menores quando comparada com a estação seca (maio a setembro) da região. Destaca-se que os valores mínimos da radiação global como outlier na estação chuvosa para a maioria das EMAs avaliadas, estão relacionados com a presença de nuvens, todavia, também podem ocorrer na estação seca, quando associados às queimadas, que são comuns na região nesse período e emitem material particulado que permanece em suspensão na atmosfera, atenuando assim a Hg por absorção e difusão não seletiva e reduzindo a transmissividade atmosférica.

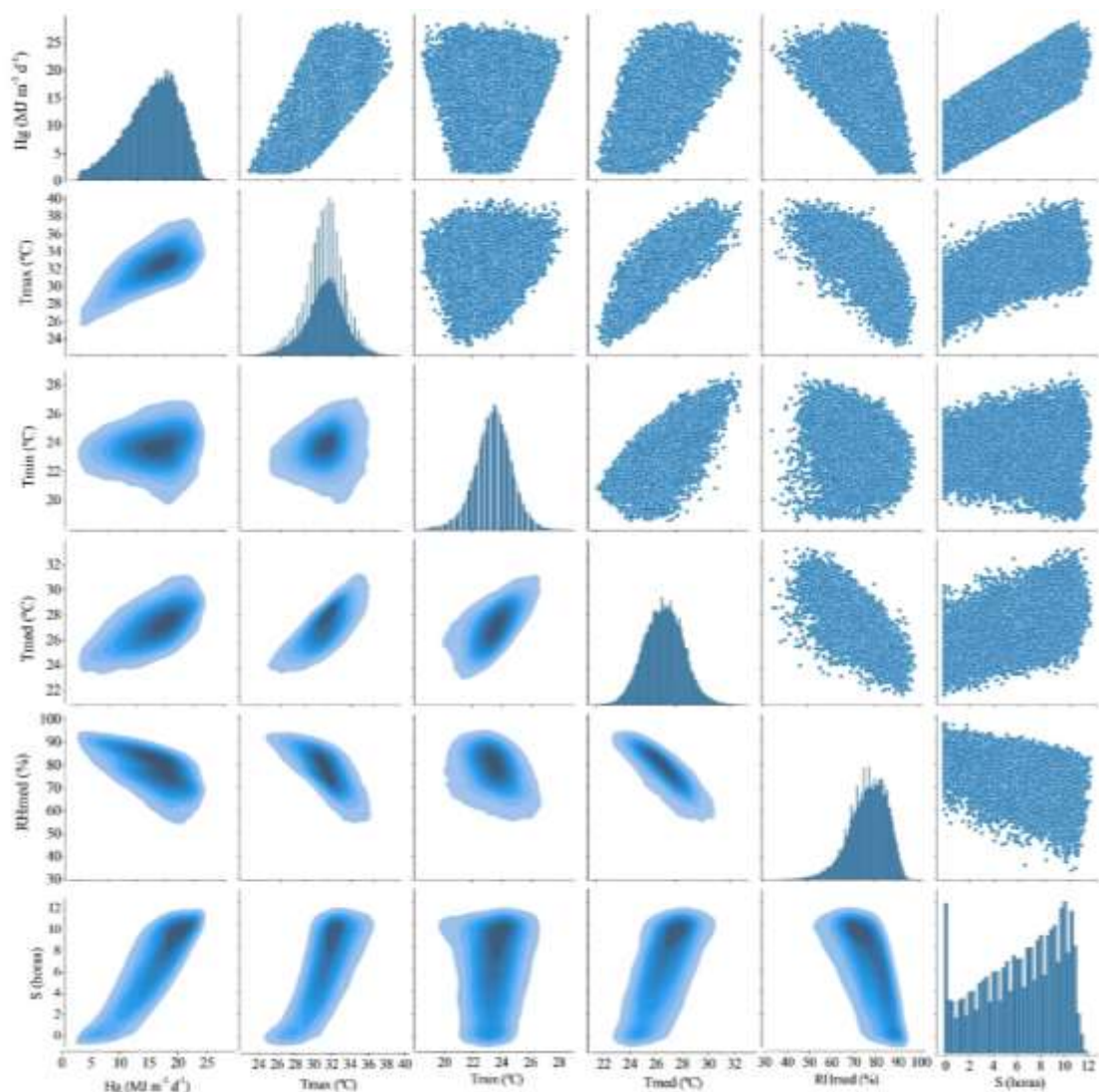


Figura 3.2. Correlações, distribuições de frequência e interpolação entre as variáveis meteorológicas (Hg, Tmax, Tmin, Tmed, RHmed, S) no bioma Amazônico brasileiro.

Outra forma de se observar as variações de Hg na Amazônia brasileira, pode ser realizada entre estações meteorológicas localizadas nas diferentes latitudes de inserção do bioma (Figura 3.4); em que, comparando a sazonalidade diária da Hg ao longo do período analisado para as estações meteorológicas de Boa Vista ($2,82^{\circ}$ N), Manaus ($3,10^{\circ}$ S) e Sinop ($11,85^{\circ}$ S), nota-se que em Sinop a Hg máxima chegou próximo dos $30 \text{ MJ m}^{-2} \text{ d}^{-1}$, e ocorre menor variação das médias diárias na estação seca, quando comparado com a estação da chuva. Continuando, como a época do ano influencia da radiação extraterrestre (H_0), têm-se que na estação seca, em função da declinação solar e da latitude de Sinop, os valores de H_0 diários são menores, e, conseqüentemente

podem gerar maiores coeficientes de transmissividade atmosférica quando comparados com outras estações meteorológicas do bioma.

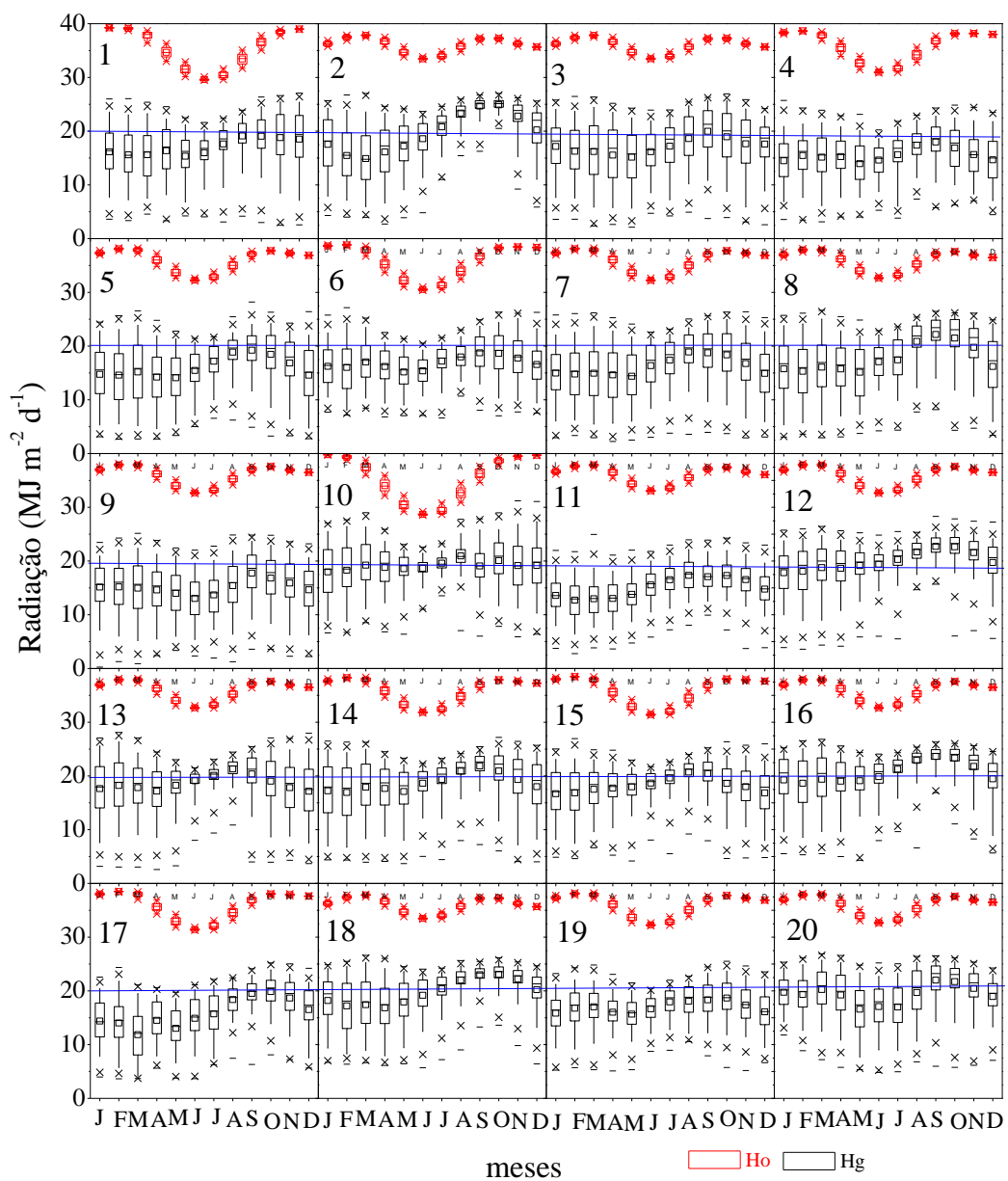


Figura 3.3. Boxplots da radiação global medida (Hg) e da radiação extraterrestre calculada (Ho) para 20 estações meteorológicas na Amazônia brasileira, em agrupamentos de dados mensal.

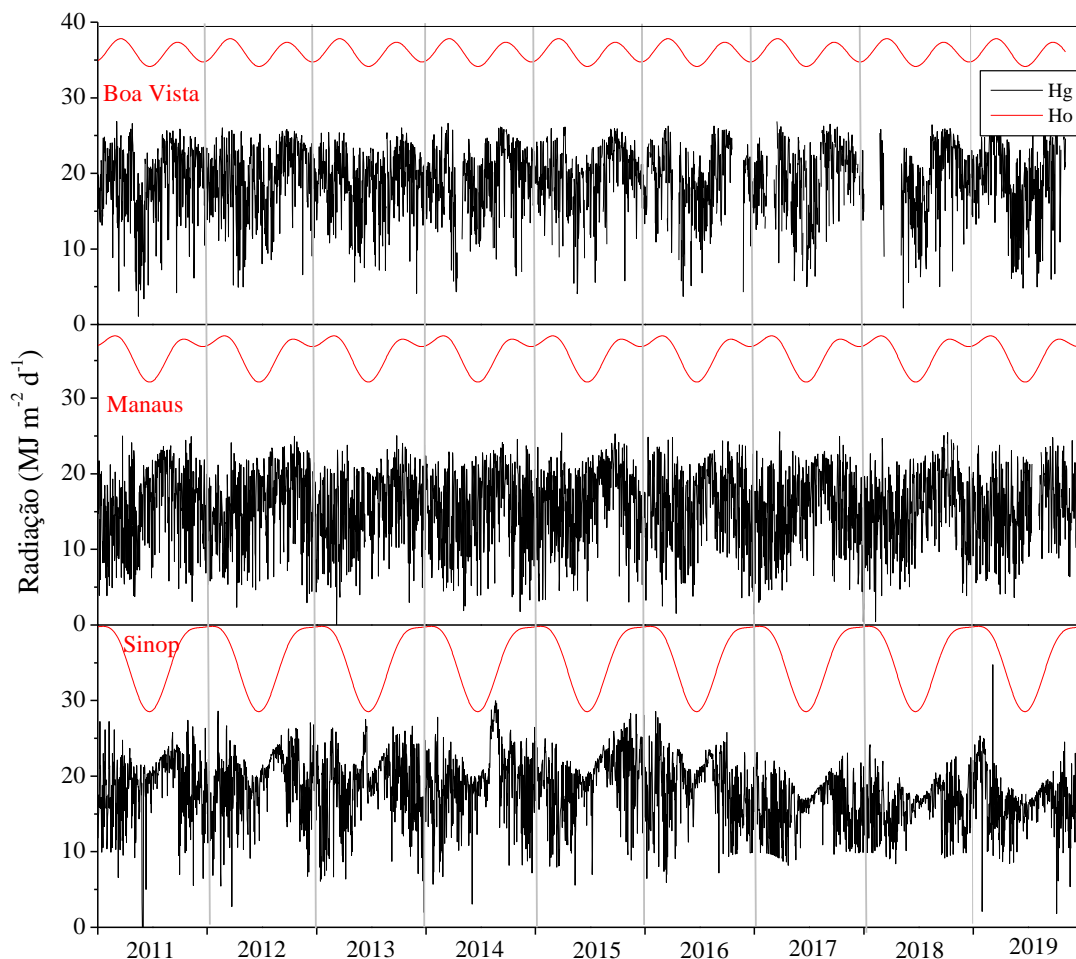


Figura 3.4. Variação da radiação global diária (Hg) e extraterrestre (Ho) para as estações meteorológicas de Boa Vista, Manaus e Sinop, na Amazônia brasileira, ao longo de diferentes anos de medidas.

Utilizando o software Statistica, foi realizada a calibração dos coeficientes específicos (a, b, c, d) de cada modelo simplificado de estimativa de Hg. E, ao se empregar o Teste-t ao nível de significância de 5%, verificou-se a significância dos coeficientes dos coeficientes ajustados para cada modelo. Os valores dos coeficientes ajustados para os melhores modelos nas 20 estações meteorológicas avaliadas, podem ser consultados na Tabela 3.7. Os resultados dos indicadores de desempenho estatístico (MBE, RMSE e “d” de Willmott) foram apresentados apenas para os modelos com coeficientes ajustados significativos a 5%.

O desempenho estatístico dos modelos empíricos foi representado em boxplots, agrupando todas as estações meteorológicas (Figura 3.5). Analisando os modelos do grupo I, observa-se ausência de melhorias significativas quando se compara as

alterações polinomiais do modelo tradicional de Angstrom-Prescott (modelo 1). No geral, para esse grupo, os valores dos coeficientes de determinação (R^2) desvios relativos (MBE), espalhamentos (RMSE) e ajustamentos (d) foram em média de 0,7780, $-0,01 \text{ MJ m}^{-2} \text{ d}^{-1}$, $2,09 \text{ MJ m}^{-2} \text{ d}^{-1}$ e 0,9310, respectivamente. Nesse grupo, o modelo 11 ($\text{Hg}/\text{Ho} = a^{1/(f(S)/S_0)}$) apresentou os piores desempenhos estatísticos, com R^2 , MBE, RMSE e d médio de 0,4980, $-0,329 \text{ MJ m}^{-2} \text{ d}^{-1}$, $3,34 \text{ MJ m}^{-2} \text{ d}^{-1}$, 0,9140, respectivamente. E, quando se compara os modelos que apresentam a mesma base analítica, observa-se que o modelo 10 [$\text{Hg}/\text{Ho} = a + b (S/S_0)^c$] apresentou desempenho estatístico superior ao restante dos modelos dependentes da insolação (S) com R^2 , MBE, RMSE e d médio de 0,7990, $0,017 \text{ MJ m}^{-2} \text{ d}^{-1}$, $1,95 \text{ MJ m}^{-2} \text{ d}^{-1}$ e 0,9400.

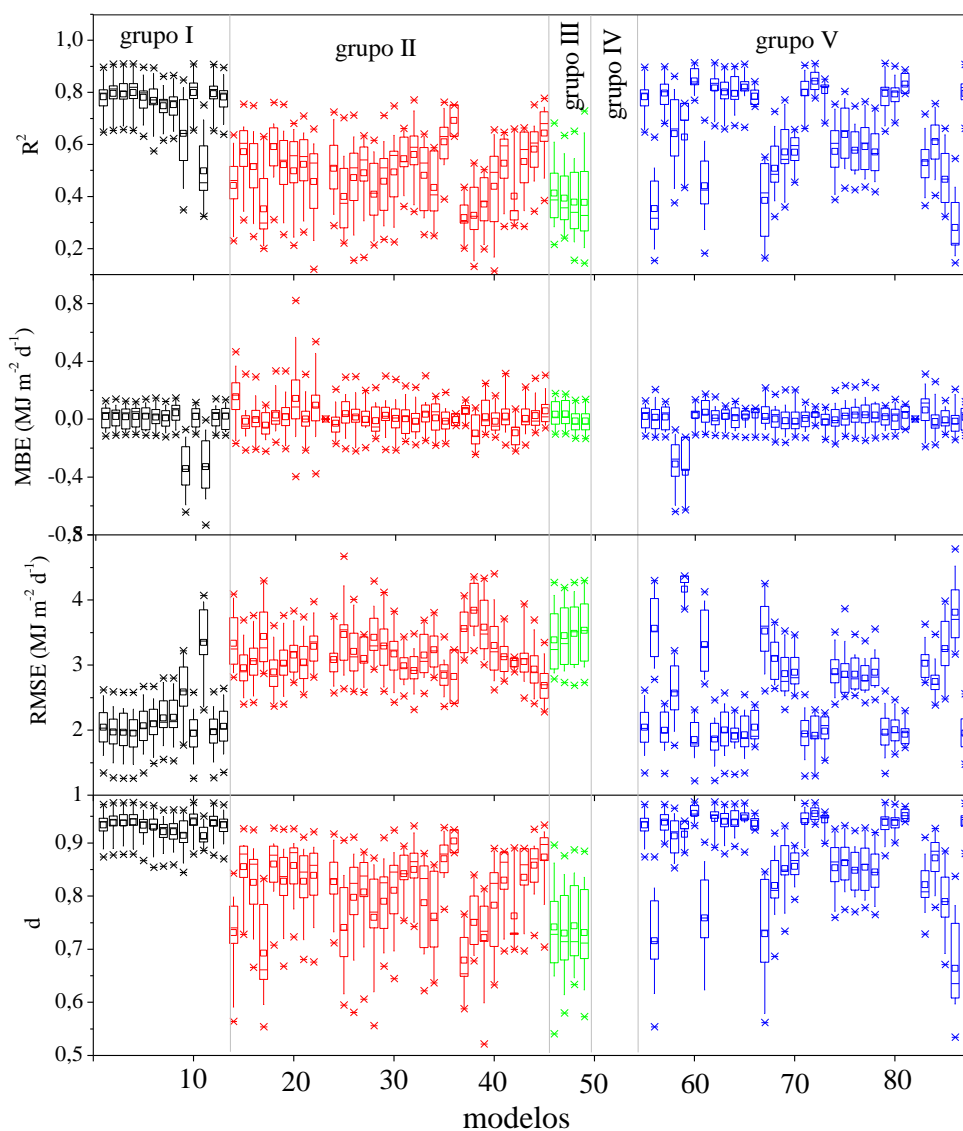


Figura 3.5. Boxplots dos indicadores de desempenho estatístico (R^2 , MBE, RMSE e d) para os 87 modelos simplificados de estimativas de Hg, considerando os valores agrupados de 20 estações meteorológicas do bioma Amazônico brasileiro. Em que: coeficiente de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE) e índice de concordância de Willmott (d).

Quanto aos modelos simplificados do grupo II (baseados na temperatura do ar), o modelo 23 [$Hg/Ho = a + b T_{min}$] dependente apenas da T_{min} e Ho foi NS para todas as estações meteorológicas avaliadas. Esse comportamento já é esperado em virtude dos instantes de ocorrência da temperatura mínima do ar, não ser concomitantes com a incidência de radiação solar. Já o modelo 24 [$Hg/Ho = a + b T_{max}$], dependente de T_{max} e Ho foi o único modelo com coeficientes significativos em todas as estações meteorológicas avaliadas, todavia, na média não apresentou os melhores indicadores de desempenho estatístico.

De forma genérica, dos 32 modelos avaliados nesse grupo (envolvendo a temperatura do ar), o modelo 17 [$Hg/Ho = a + b/Ho + c T_{med}/Ho$] apresentou o pior desempenho estatístico, com R^2 , MBE, RMSE e d médio de 0,3520, 0,099 $MJ m^{-2} d^{-1}$, 3,43 $MJ m^{-2} d^{-1}$ e 0,6930, respectivamente; já o modelo 45 [$Hg/Ho = a + b \Delta T + c \Delta T^{0,25} + d \Delta T^{0,5} + e(T_{med}/Ho)$] apresentou o melhor desempenho com R^2 , MBE, RMSE e d médio de 0,6430, 0,056 $MJ m^{-2} d^{-1}$, 2,68 $MJ m^{-2} d^{-1}$ e 0,8730, todavia, não apresentando ajustes significativos para as estações meteorológicas localizadas em Cameté e Tucuruí. Normalmente, os modelos empíricos que empregam a amplitude térmica ($\Delta T = T_{max} - T_{min}$) como variável de entrada apresentaram melhores estimativas. Comparando o desempenho dos modelos 14 [$Hg/Ho = a \Delta T^{0,5}$] e modelo 16 [$Hg/Ho = a + b \Delta T^{0,5}$], observa-se que a inserção do coeficiente linear, melhora o desempenho estatístico significativamente, visto que os valores de R^2 variaram de 0,4420 para 0,5150, com redução de MBE de 0,154 para -0,003 $MJ m^{-2} d^{-1}$ e de RMSE de 3,33 para 3,05 $MJ m^{-2} d^{-1}$, e ainda, com aumento do índice de ajustamento (d) de 0,7312 para 0,8257.

Os modelos simplificados do grupo III (baseados apenas na umidade relativa média diária - RHmed), não apresentaram coeficientes significativos pelo Teste-t nas estações meteorológicas de Eirunepé, São Gabriel da Cachoeira e Sinop. Em regra, o desempenho estatístico dos modelos do grupo III foram os piores quando comparados

aos modelos dos outros grupos, sendo que, nesse caso, as melhores estimativas desse grupo foram geradas pelo modelo 46 [$Hg/Ho = a / Ho + b(RHmed/Ho)$], com valores de R^2 , MBE, RMSE e d de 0,4130, 0,034 $MJ m^{-2} d^{-1}$, 3,38 $MJ m^{-2} d^{-1}$ e 0,7420, respectivamente. Quanto aos modelos do grupo IV (baseados em variáveis astronômicas) todos não apresentaram significância estatística, para nenhuma das estações meteorológicas avaliadas. No grupo V (modelos híbridos), as melhores estimativas de Hg foram geradas pelo modelo 62 [$Hg/Ho = a + b \ln\Delta T + c (S/So)^d$], com valores médios de R^2 , MBE, RMSE e d de 0,8170, 0,008 $MJ m^{-2} d^{-1}$, 1,86 $MJ m^{-2} d^{-1}$ e 0,9460, respectivamente.

Considerando apenas os modelos que apresentaram os melhores desempenhos estatísticos em cada grupo (Figura 3.6), observa-se que na maioria das estações meteorológicas, as melhores estimativas de Hg foram geradas pelos modelos 62 e 10. Nesse caso, o modelo 62 apresentou valores que variam de 0,6714 a 0,9137 para R^2 , de -0,112 a 0,154 $MJ m^{-2} d^{-1}$ para MBE, de 1,22 a 2,46 $MJ m^{-2} d^{-1}$ para RMSE, e, de 0,8921 a 0,9772 para o índice de ajustamento (d); já o modelo 10, apresentou variações de R^2 de 0,6549 a 0,9091, de MBE de -0,103 a 0,115 $MJ m^{-2} d^{-1}$, de RMSE de 1,26 a 2,58 $MJ m^{-2} d^{-1}$ e índice “ d ” de 0,8793 a 0,9760. A diferença principal entre as variáveis de entrada desses dois modelos, é a adição do logaritmo neperiano (\ln) na amplitude térmica ($T_{max} - T_{min}$) no modelo 62, que não propiciou melhorias significativas na capacidade preditiva do modelo. Todavia, ressalta-se que esses dois modelos (62 e 10) são baseados na insolação, dessa forma, em casos de disponibilidade mínima de dados, preferencialmente deve ser empregado o modelo 45 que depende apenas da temperatura do ar, visto que este gera melhores respostas do que os modelos associados a umidade relativa do ar.

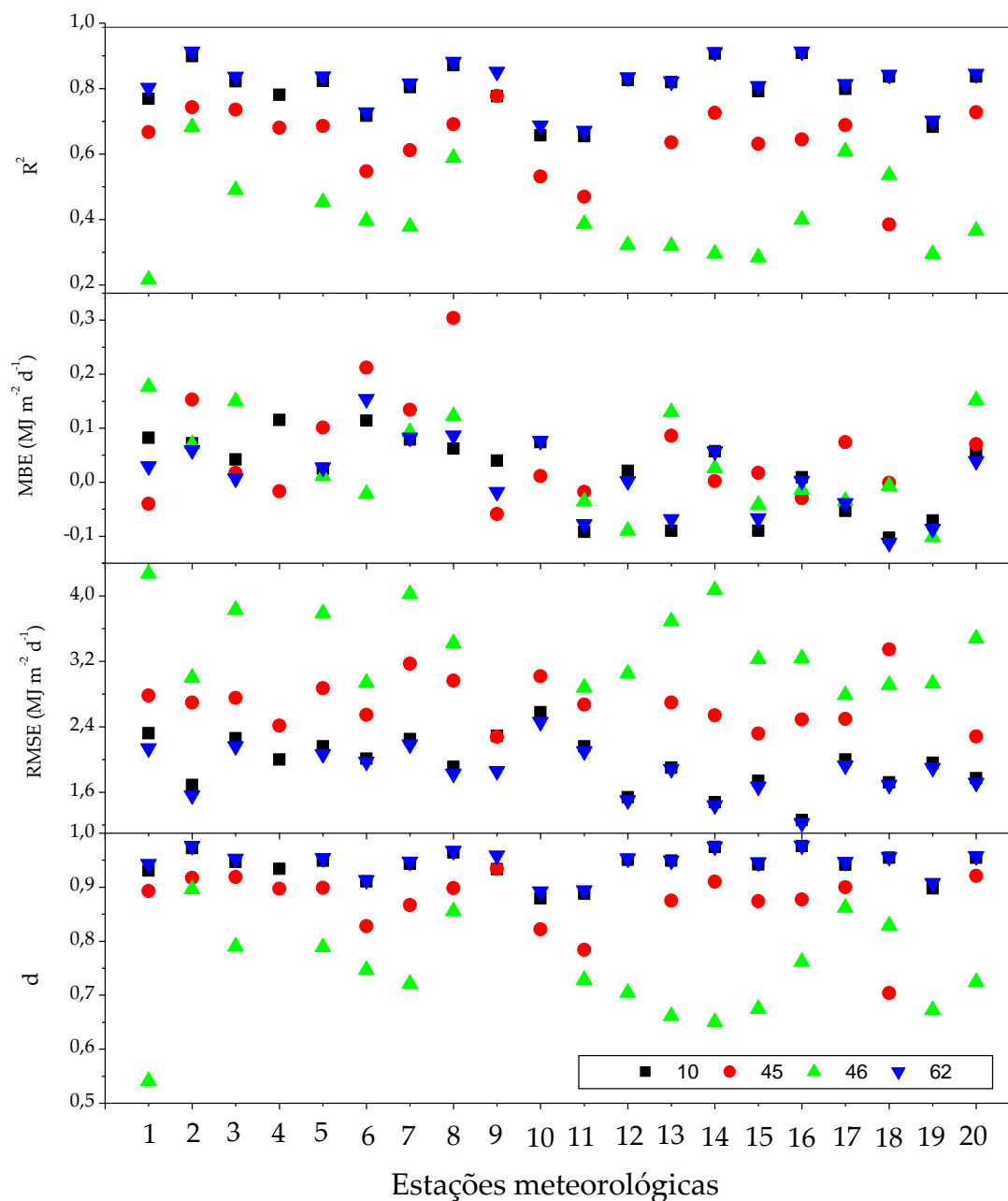


Figura 3.6. Indicativos de desempenho estatístico dos melhores modelos simplificados de estimativas de cada grupo (modelos 10, 45, 46 e 62) para 20 estações meteorológicas do bioma Amazônico brasileiro. Em que coeficiente de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE) e índice de concordância de Willmott (d).

Na Figura 3.7 são apresentadas as dispersões de Hg medida e estimada pelos modelos 10, 45 e 62, para as três estações meteorológicas supracitadas que representam a variação de latitude no bioma Amazônico (Boa Vista, Manaus e Sinop).

Nesse caso, foram consideradas quatro diferentes classes de transmissividade atmosférica, que representam as condições de céu nublado ($0 < Kt < 0,35$), parcialmente nublado com predominância de radiação difusa ($0,35 < Kt < 0,55$), parcialmente aberto com predominância de radiação direta ($0,55 < Kt < 0,65$) e céu aberto ($Kt > 0,65$), conforme recomendações de Escobedo et al. (2009) e Souza et al. (2016). A radiação global estimada pelos modelos 10 e 62 são semelhantes e seguem próxima da linha de referência (1:1), com maior dispersão dos valores estimados quando o Kt se encontra abaixo dos 0,55.

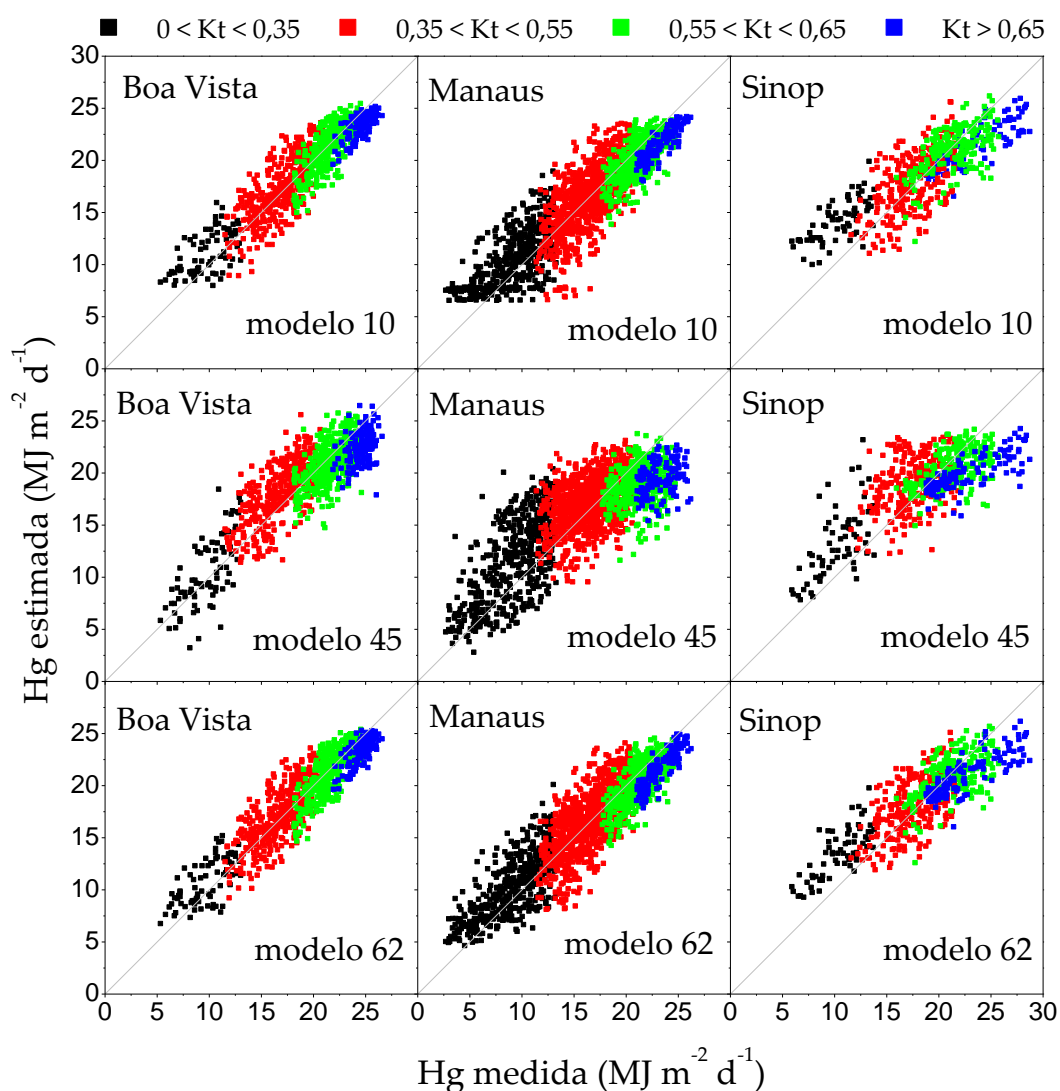


Figura 3.7. Dispersão entre a radiação global medida e estimada pelos modelos 10, 45 e 62 para as estações meteorológicas de Boa Vista (RR), Manaus (AM) e Sinop (MT),

em diferentes classes de transmissividade atmosférica. (A linha cinza, representa a relação 1:1 ou $y = x$).

O erro absoluto acumulado na frequência de ocorrência (Figura 3.8) até o valor de $2,0 \text{ MJ m}^{-2} \text{ d}^{-1}$ na estimativa da Hg empregando os modelos 10, 45 e 62 nas estações meteorológicas de Boa Vista foi de 84, 83, 85% e Sinop foi de 67, 68 e 68%, porém em Manaus houve diferença expressiva dos valores, com 75, 64 e 82%. Dependendo da cidade avaliada, a frequência do erro acumulado até $2,0 \text{ MJ m}^{-2} \text{ d}^{-1}$ pode ser alta (Boa Vista) ou baixa (Sinop), ou seja, o erro é diretamente influenciado pelas condições meteorológicas locais. A redução do erro absoluto acumulado em até $2,0 \text{ MJ m}^{-2} \text{ d}^{-1}$ nas condições meteorológicas de baixa transmissividade atmosférica ($0 < Kt < 0,35$), é mais eficiente no modelo híbrido 62 (55 a 85%) quando comparado ao modelo 10 (33 a 80%) e principalmente o modelo 45 (38 a 74%), seguindo a mesma tendência em condições meteorológicas de alta transmissividade atmosférica ($Kt > 0,65$) com frequência acumulada do modelo 62 de 82 a 93%, modelo 10 de 78 a 87% e modelo 45 de 32 a 71% respectivamente. Na condição de céu parcialmente nublado com predominância de radiação difusa ($0,35 < Kt < 0,55$), quando se compara o modelo 10 (67 a 81%) e modelo 62 (67 a 79%), não houve diferença no erro acumulado, porém o modelo 45 (60 a 70%) a maior possibilidade de erros acima dos $2,0 \text{ MJ m}^{-2} \text{ d}^{-1}$, sendo que esse padrão se repetiu na condição de céu parcialmente aberto com predominância de radiação direta ($0,55 < Kt < 0,65$) com modelo 10 (87 a 93%), modelo 62 (88 a 92%) e modelo 45 (71 a 84%).

Para permitir aplicações e estimativas de Hg, são apresentados os coeficientes ajustados apenas para esses três modelos simplificados (10, 45 e 62), nas 20 estações meteorológicas avaliadas no bioma Amazônia (Tabela 3.7). Em algumas cidades, como observado na Tabela 3.7, os modelos 45 e 62 foram NS, sendo que, para o grupo II (baseados na temperatura do ar), recomenda-se o modelo 65 para a cidade de Cametá [$\text{Hg} = (-0,0942 + (-0,2688 + 0,0179 * T_{\text{med}}) * \Delta T^{0,5}) * H_0$] e Tucuruí [$\text{Hg} = (0,0655 + (-0,1478 + 0,0108 * T_{\text{med}}) * \Delta T^{0,5}) * H_0$] e no grupo V (combinação híbrida) o modelo 72 apresentou o melhor desempenho para a cidade de Eirunepé [$\text{Hg} = (-0,0947 - 0,0077 * T_{\text{min}} + 0,0188 * T_{\text{max}} + 0,3163 * (S/S_0)) * H_0$].

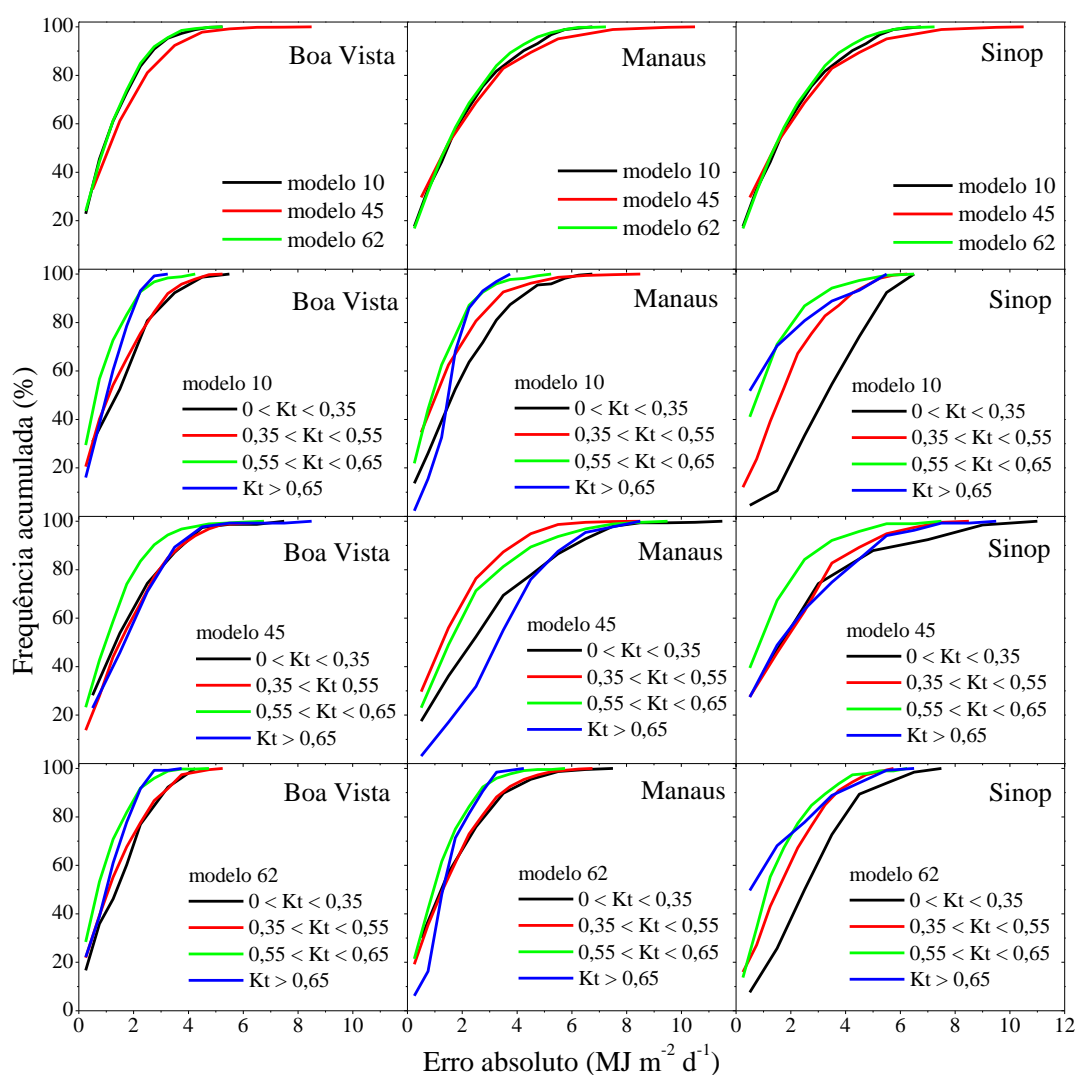


Figura 3.8. Frequência do erro absoluto nas estações meteorológicas de Boa Vista, Manaus e Sinop, para os modelos 10, 45 e 62 em diferentes transmissividades atmosférica.

Tabela 3.7. Coeficientes calibrados para os modelos 10, 45 e 62 de estimativas da radiação global diária, para cada estação meteorológica avaliada no Bioma Amazônico.

Estação	modelo 10				modelo 62				
	a	b	c	R ²	a	b	c	d	R ²
1	0,2122*	0,4784*	0,7003*	0,772	0,0324*	0,1163*	0,3607*	0,8035*	0,8052
2	0,1995*	0,538*	0,7178*	0,8958	0,0396*	0,1123*	0,4516*	0,8104*	0,9071
3	0,2142*	0,4846*	0,5639*	0,8045	0,0351*	0,1127*	0,3832*	0,6435*	0,8262
4	0,2233*	0,4343*	0,5737*	0,758	0,0013 ^{ns}	0,1350 ^{ns}	0,3097*	0,6808 ^{ns}	0,7979
5	0,1817*	0,4678*	0,6911*	0,8221	0,0479*	0,0936*	0,3885*	0,7413*	0,8376
6	0,2769*	0,3842*	0,6852*	0,7606	0,1265*	0,0815*	0,3178*	0,6830*	0,7777

7	0,2041*	0,4737*	0,7516*	0,7846	0,0518*	0,0969*	0,4094*	0,8466*	0,7992
8	0,1672*	0,5270*	0,6562*	0,8933	0,0733*	0,0728*	0,4662*	0,7124*	0,9003
9	0,1825*	0,4718*	0,6701*	0,7713	-0,0453*	0,1738*	0,2741*	0,9689*	0,8673
10	0,3103*	0,4191*	0,8520*	0,6359	0,1515*	0,0821*	0,3388*	0,8233*	0,665
11	0,2171*	0,3681*	0,9748*	0,6563	0,0794*	0,0846*	0,3203*	1,25*	0,6658
12	0,2159*	0,4893*	0,7101*	0,8452	0,1109*	0,0687*	0,4474*	0,7844*	0,852
13	0,1970*	0,4940*	0,6848*	0,8221	0,1327*	0,0410*	0,4448*	0,6997*	0,8263
14	0,2113*	0,4940*	0,6286*	0,887	0,1042*	0,0718*	0,4294*	0,6683*	0,8954
15	0,2355*	0,4382*	0,6629*	0,8081	0,0935*	0,0862*	0,3600*	0,6808*	0,8256
16	0,2055*	0,5173*	0,6947*	0,9029	0,1331*	0,0499*	0,4823*	0,7233*	0,9068
17	0,1968*	0,4229*	0,7587*	0,794	0,0487*	0,0876*	0,3483*	0,8363*	0,8065
18	0,278*	0,4029*	0,6001*	0,8422	0,2053*	0,0450*	0,4053*	0,6424*	0,8477
19	0,2549*	0,3692*	0,7780*	0,6905	0,1230*	0,0836*	0,3203*	0,9069*	0,7108
20	0,2453*	0,4656*	0,7138*	0,8429	0,1176*	0,0766*	0,4036*	0,7679*	0,8503

Estação	modelo 45					R ²
	a	b	c	d	e	
1	7,10*	-0,4130*	-13,02*	6,47*	0,3146*	0,6513
2	18,05*	-1,06*	-32,94*	16,11*	1,08*	0,7264
3	6,68*	-0,3971*	-13,03*	6,39*	1,04*	0,7197
4	4,56*	-0,2595*	-8,59*	4,26*	0,4144*	0,6825
5	12,48*	-0,7875*	-23,28*	11,62*	0,7932*	0,7007
6	10,72*	-0,5403*	-18,54*	8,83*	0,3082*	0,6095
7	14,57*	-0,8666*	-26,66*	13,08*	0,8747*	0,6124
8	13,14*	-0,9469*	-25,69*	13,19*	1,17*	0,7341
9	1,52*	-0,0997*	-3,45*	1,84*	0,4104*	0,8000
10	7,33*	-0,3692*	-12,76*	6,12*	0,2696*	0,4998
11	13,04*	-0,6702	-22,98*	10,87*	0,8086*	0,4661
12	2,79 ^{ns}	-0,3656*	-7,70 ^{ns}	4,57*	0,9133*	0,5068
13	2,93*	-0,2000*	-6,02*	3,16*	0,3889*	0,629
14	13,70*	-0,8282*	-25,04*	12,42*	0,5538*	0,6981
15	5,06*	-0,3339*	-9,95*	5,10*	0,4996*	0,6639
16	27,61*	-1,71*	-50,07*	24,76*	1,09*	0,6741
17	13,64*	-0,6925*	-23,86*	11,31*	0,5745*	0,6767
18	-33,41*	1,81*	55,84*	-26,79*	1,72*	0,3738
19	-1,10 ^{ns}	-0,0433 ^{ns}	0,2451 ^{ns}	0,3368 ^{ns}	0,7507**	0,4570
20	11,07*	-0,5971*	-20,36*	9,72*	1,17*	0,7435

3.4 DISCUSSÃO

Os resultados encontrados neste estudo fornecem detalhes da calibração local e significância estatística dos coeficientes de 87 modelos de estimativa diária da

radiação global em regiões de clima tropical, tendo como variáveis de entrada, dados meteorológicos que podem estar facilmente disponibilizados por medições de rotina em estações meteorológicas. A determinação da significância dos coeficientes de cada modelo é uma informação relevante na modelagem, pois quanto menor o P-valor, do Teste-t ao nível de significância α (5%), melhor é a capacidade preditiva do modelo, e, quando existe coeficientes não significativos (NS), indica que na região avaliada, as condições meteorológicas e ambientais podem ser complexas e causarem maiores incertezas na estimativa de Hg (BAKIRCI, 2009). Isso pode ser observado pelo grande número de modelos com coeficientes calibrados localmente que foram NS (Figura 3.9 em anexo).

Como observado nas Figuras 3.2 a 3.4, a insolação (S) e a radiação global (Hg) incidente no bioma Amazônico, são significativamente influenciadas pelos componentes atmosféricos, como por exemplo a emissão de aerossóis para a atmosfera através da queima da vegetação na estação seca da bacia Amazônica (SILVA JUNIOR et al. 2022) ou com a presença de nuvens em razão da alta precipitação pluvial na estação da chuva (ALVARES et al. 2013; ARAÚJO et al. 2024). Sendo essas condições o principal fator de atenuação da Hg, e o aumento da incerteza nos valores de insolação, em função da sensibilidade à queima dos heliogramas e da manutenção dos heliógrafos. De acordo com os resultados (Figura 3.5), o melhor desempenho estatístico foi obtido com a modificação do modelo linear de Angstrom-Prescott para um modelo que faz uso de potência em sua variável explicativa $[Hg/Ho = a + b (S/So)^c]$. Segundo Almorox & Hontoria (2004) e Bakirci (2009), dependendo das condições de transmissividade atmosférica, a relação entre S e Hg pode ser melhor representada por correlações do tipo linear, polinomial, logarítmica, exponencial ou híbrido. Para Santos et al. (2021), mesmo em condições de ausência de nebulosidade (céu aberto), elementos presentes na atmosfera terrestre podem atenuar a radiação solar por difusão (espalhamentos) e influenciar a Hg.

Avaliando os diferentes modelos do grupo I, nas condições climáticas da Turquia, Bakirci (2009) não observou diferenças expressivas nos desempenhos estatísticos dos modelos linear, exponencial e logarítmico, obtendo coeficientes de determinação (R^2) superiores a 0,96 e valores de MBE que variaram de 0,20 a 0,24 MJ m⁻² d⁻¹ e de RMSE de 1,31 a 1,35. Já Almorox & Hontoria (2004), em dezesseis cidades da Espanha,

avaliaram modelos linear e polinomiais de segundo e terceiro grau, logarítmicos e exponenciais baseados na insolação (S) e concluíram que todos os modelos estatísticos podem ser utilizados na estimativa de Hg com boa precisão, porém, recomendam o modelo linear pela simplicidade e por apresentar melhores desempenhos estatísticos.

Observou-se que o desempenho dos modelos 8, 9 e 11 que não possuem coeficiente linear (a), foi menor em todas as cidades. Segundo Prieto & García (2022) esse coeficiente representa a transmissividade mínima da atmosfera da região, e está associado a radiação difusa. Na Amazônia, essa condição representa parte considerável da radiação global em função da sazonalidade das precipitações pluviais e alterações na composição da atmosfera (SOUZA et al. 2016), e, portanto, modelos que não apresentem coeficientes que caracterizem dias com baixa transmissividade atmosférica, podem gerar erros maiores para ajustes em agrupamentos anuais ou sazonais; em casos de ajustes mensais, esses modelos potencialmente podem apresentar boas estimativas para os meses secos (ausência de nebulosidade) em algumas regiões da Amazônia.

Ainda no tocante as inferências, em algumas regiões, as mudanças climáticas podem aumentar a precipitação, que resulta na redução na transmissividade atmosférica e conseqüentemente nos ajustes dos modelos, como é o caso que ocorre na cidade de Boa Vista, segundo Araújo et al. (2024), analisando o período de dados meteorológicas de 1961 a 2020, houve aumento da precipitação de 1.420,4 para 1.761,8 mm ano⁻¹ e na temperatura média do ar de 27,4 para 28,2 °C.

Para os modelos do grupo II (baseados na temperatura do ar), as melhores estimativas da radiação global diária, foram obtidas com o modelo 45 [$Hg/Ho = a + b (T_{max} - T_{min}) + c (T_{max} - T_{min})^{0,25} + d (T_{max} - T_{min})^{0,50} + e (T_{med}/Ho)$], corroborando com os resultados encontrado por Qiu et al. (2022). Esses autores, avaliaram 78 modelos baseados na temperatura do ar para 105 estações meteorológicas na China, e concluíram que modelos que relacionam amplitude térmica ($\Delta T = T_{max} - T_{min}$), temperatura máxima, média e mínima (T_{max} , T_{med} e T_{min}) apresentam melhores estimativas. Nos modelos baseados na temperatura do ar, a ΔT está relacionada a diversos fatores locais que influenciam diretamente no balanço de radiação e de energia, como latitude, precipitação pluvial, pressão atual de vapor d'água, transmissividade atmosférica (nebulosidade), proximidade de grandes

superfícies de água livre, dentre outros fatores (SANTOS et al. 2021). Esse é um parâmetro essencial nos modelos baseados somente na temperatura do ar, quando se deseja melhorar a capacidade preditiva, pois ΔT possui uma boa correlação com a Hg (QIU et al. 2022).

Destaca-se ainda que o desempenho de modelos empíricos de estimativas da Hg pode ser considerado satisfatório ou insatisfatório, dependendo da região de estudo e das variáveis de entrada, como por exemplo, neste estudo em condições tropicais, o modelo 37 [$Hg/Ho = a + b Ho T_{med}^c$] apresentou o pior desempenho no grupo II com coeficiente de determinação de 0,2765 a 0,4000 e os coeficientes ajustados foram não significativos em 55% das cidades avaliadas, porém para a região do Cairo (Egito), clima seco e desértico, Hassan et al. (2016), avaliando 20 modelos baseados na temperatura do ar, verificaram que este mesmo modelo apresentou os melhores desempenhos estatísticos com RMSE de $0,5813 MJ m^{-2} d^{-1}$ e R^2 de 0,9897.

Pode-se considerar que a significância e também o desempenho dos modelos empíricos calibrados regionalmente é a junção de dois fatores, sendo o primeiro relativo as variáveis meteorológicas de entrada que representam as condições ambientais do local, e o segundo devido a estrutura do modelo empírico.

Os menores desempenhos na estimativa de Hg foram obtidos com os modelos empíricos baseados na umidade relativa do ar (grupo III). Modelos do grupo III, geralmente apresentaram boas estimativas em regiões áridas com baixos totais anuais de precipitação pluvial como Nigéria (FALAYI et al. 2008; KOLEGABE et al. 2016), Barém (ELAGIB et al. 1998) e Turquia (ERTEKIN & YALDIZ, 1999).

Os cinco modelos do grupo IV foram NS para todas as 20 estações meteorológicas avaliadas (Figura 3.9 em anexo). Variáveis de entrada astronômicas, como a declinação solar, fotoperíodo e dia juliano são variáveis que não ponderam os fatores locais, relacionados as condições atmosféricas e geográficas, que podem interferir diretamente na radiação global incidente na superfície; onde a utilização de frações radiométricas objetiva, justamente minimizar esses efeitos (ESCOBEDO et al. 2009; SOUZA et al. 2016). Os modelos do grupo IV são empregados em países com regiões áridas, baixa precipitação pluvial e alta transmissividade atmosférica como a Arábia Saudita (BENGHANEM & MELLIT, 2014) e Jordânia (AL-SALAYMEH, 2006).

Embora os desempenhos estatísticos de modelos híbridos (grupo V) tenham sido melhores quando comparado aos modelos simples com temperatura do ar (grupo II), umidade relativa (grupo III), essas diferenças são reduzidas quando comparado com modelos baseados na insolação (grupo I) (Figura 3.6). Vários trabalhos relataram diferenças consideráveis no desempenho de modelos simples em comparação aos modelos híbridos e em diferentes regiões climáticas (CHEN & LI, 2013; JAHANI et al., 2017; FAN et al., 2018; SAMANTA et al., 2021; SANTOS et al., 2021; PIETRO & GARCIA, 2022).

Quando se analisa as variações das estimativas em estações meteorológicas de diferentes latitudes na Amazônia (Figura 3.7), observa-se novamente que em geral, os modelos apresentam maiores dificuldades de estimativas em dias e épocas do ano com transmissividade atmosférica intermediária (céu parcialmente nublado ou parcialmente aberto). Segundo Santos et al. (2021), diferentes combinações de nebulosidade podem resultar no mesmo valor de ΔT , porém cada combinação de nebulosidade resulta em um único valor da Hg.

Outro ponto a ser destacado, é que a utilização de modelos baseados na temperatura do ar, geralmente propicia a ocorrência de superestimativas de Hg (ABRAHA & SAVAGE, 2008; FAN et al. 2018; SAMANTA et al. 2021; SANTOS et al. 2021). Esse comportamento ocorre, pois, a temperatura do ar é dependente de processos de aquecimento da superfície, do balanço de energia e do transporte de calor, que por sua vez, pode refletir em níveis de calor sensíveis superiores aos que potencialmente seriam gerados exclusivamente pela Hg incidente.

Os modelos 10, 45 e 62 apresentaram consistências nas estimativas de Hg, cada um com variáveis de entrada próprias (grupos) e podem ser empregados na Amazônia brasileira em função da distribuição espacial das 20 estações meteorológicas avaliadas.

Apesar dos bons ajustes de modelos empíricos encontrados neste trabalho, segundo Prieto & García (2022), após revisarem artigos publicados nos últimos 100 anos com essa temática, verificaram que nenhum dos 165 modelos empíricos com diferentes variáveis de entrada foram consistentes e apresentaram bons desempenhos nas diferentes regiões climáticas. Isso se torna mais preocupante, com cenários de mudanças climáticas, que preveem a elevação da temperatura do ar, mudança nos padrões de chuvas e alteração na concentração de gases, que conseqüentemente, irão

interferir diretamente no balanço de radiação (SAMANTA et al. 2021; DELGADO et al. 2022). Esse cenário de mudanças climáticas, vêm sendo observado também na região do bioma Amazônico, com aumento da precipitação pluvial e temperatura do ar (ARAÚJO et al. 2024; SABINO et al. 2024), portanto, torna-se fundamental entender e disponibilizar o maior número possível de modelos de estimativas de variáveis meteorológicas, que permitam calibrações regionais e aplicações em grandes bancos de dados, para preenchimentos adequados de falhas e geração de bases de dados longas e consistentes. E, nesse caso, quando essas condições de mudanças climáticas se concretizarem, os modelos empíricos devem ser recalibrados e surgirão novos estudos, principalmente em agrupamentos de dados menores (sazonais e mensais), devem ser conduzidos para atualizar os coeficientes e manter a capacidade preditiva dos diferentes modelos empíricos.

3.5 CONCLUSÃO

Os modelos empíricos mais precisos para estimar a radiação global (Hg) para o bioma da Amazônia brasileira foram primeiro o modelo híbrido baseado na insolação e na temperatura do ar, representado por $[Hg/Ho = a + b \ln\Delta T + c S/So^d]$, seguido pelo modelo baseado na insolação $[Hg/Ho = a + b S/So^c]$, e finalmente o modelo baseado na temperatura do ar $[Hg/Ho = a + b\Delta T + c\Delta T^{0,25} + d\Delta T^{0,5} + e Tmed/Ho]$ no bioma da Amazônia brasileira.

Não é recomendado o uso de modelos simplificados baseados apenas na umidade relativa ou variáveis astronômicas para estimativa de Hg na Amazônia, em agrupamentos de dados anuais; esses modelos podem apresentar melhores desempenhos estatísticos se forem calibrados localmente para estações meteorológicas com estações secas e chuvosas bem definidas.

A nebulosidade e a sazonalidade das chuvas afetam a transmissividade atmosférica e as estimativas da radiação global. Neste caso, futuras atualizações destas análises deverão considerar os modelos que geraram as melhores estimativas e buscar análises em sazonalidade (ano ou estações hídricas). e agrupamentos mensais de bancos de dados. A recomendação de modelos simplificados que produzam boas

estimativas de Hg também permitirão comparações com outras metodologias, como aprendizado de máquina e reanálise.

Portanto, a geração de condições que permitam o conhecimento da radiação global contribuirá para inúmeras aplicações agrícolas e ambientais, fornecendo um contexto abrangente para a aplicação destas estimativas em escalas espaciais e temporais maiores.

3.6 REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHA, M. G.; SAVAGE, M. J. Comparison of estimates of daily solar radiation from air temperature range for application in crop simulations. **Agriculture and Forest Meteorology**, v. 148, n. 3, p. 401-416, 2008. <https://doi.org/10.1016/j.agrformet.2007.10.001>.

ADARAMOLA, M. S. Estimating global solar radiation using common meteorological data in Akure, Nigeria. **Renewable Energy**, v. 47, p. 38-4, 2012. <https://doi.org/10.1016/j.renene.2012.04.005>.

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. Crop evapotranspiration guidelines for computing crop water requirements e FAO irrigation and drainage paper 56. Rome: Food and Agriculture Organization of the United Nations; 1998.

ALMOROX, J.; HONTORIA, C. Global solar radiation estimation using sunshine duration in Spain. **Energy Conversion and Management**, v. 45, n. 9-10, p. 1529-1535, 2004. <https://doi.org/10.1016/j.enconman.2003.08.022>.

AL-SALAYMEH, A. Modeling of global daily solar radiation on horizontal surfaces for amman city. **Emirates Journalfor Engineerring Research**, v. 11, n. 1, p. 49-56, 2006.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. de M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013. <https://doi.org/10.1127/0941-2948/2013/0507>.

ANGSTROM, A. Solar and terrestrial radiation. **Quarterly Journal of the Royal Meteorological Society**, v. 50, p. 121-5, 1924. <https://doi.org/10.1002/qj.49705021008>.

ARAÚJO, W. F.; NETO, J. L. L. M.; SANDER, C.; ALBUQUERQUE, J. A. A.; VIANA, T. V. A.; VALERO, M. A. M. Atualização da classificação climática de Boa Vista, Roraima, Brasil. **Nativa**, v. 12, n. 2, p. 236-240, 2024. DOI: <https://doi.org/10.31413/nativa.v12i2.16202>.

ARÉVALO, S. M. M.; DELGADO, R. C.; LINDEMANN, D. S.; GELSLEICHTER, Y. A.; PEREIRA, M. G.; RODRIGUES, R. A.; JUSTINO, F. B.; WANDERLEY, H. S.; ZONTA, E.; SANTANA, R. O.; SOUZA, R. S. Past and Future Responses of Soil Water to Climate Change in Tropical and Subtropical Rainforest Systems in South America. **Atmosphere**, v. 14, n. 4, p. 755, 2023. <https://doi.org/10.3390/atmos14040755>.

BADESCU, V. Assessing the performance of solar radiation computing models and model selection procedures. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 105, p. 119-134, 2013. <https://doi.org/10.1016/j.jastp.2013.09.004>.

BAHEL, V.; BAKHSH, H.; SRINIVASAN, R. A correlation for estimation of global solar radiation. **Energy**, v. 12, n. 2, p.131-135, 1987. [https://doi.org/10.1016/0360-5442\(87\)90117-4](https://doi.org/10.1016/0360-5442(87)90117-4).

BAKIRCI, K. Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. **Energy**, v. 34, n. 4, p. 485-501, 2009. <https://doi.org/10.1016/j.energy.2009.02.005>.

BENDER, F. D.; SENTELHAS, P. C. Solar Radiation Models and Gridded Database to Fill Gaps in Weather Series and to Project Climate Change in Brazil. **Advances in Meteorology**, v. 2018, n. 1, e 6204382, 2018. <https://doi.org/10.1155/2018/6204382>.

BENGHANEM, M.; MELLIT, A. A simplified calibrated model for estimating daily global solar radiation in Madinah, Saudi Arabia. **Theoretical and Applied Climatology**, v. 115, n. 1, p. 197-205, 2014. <https://doi.org/10.1007/s00704-013-0884-2>.

BRISTOW, K. L.; CAMPBELL, G. S. On the relationship between incoming solar radiation and daily maximum and minimum temperature. **Agricultural and Forest Meteorology**, v. 31, n. 2, p. 159-166, 1984. [https://doi.org/10.1016/0168-1923\(84\)90017-0](https://doi.org/10.1016/0168-1923(84)90017-0).

CHEN, R.; ERSI, K.; YANG, J.; LU, S.; ZHAO, W. Validation of five global radiation models with measured daily data in China. **Energy Conversion and Management**, v. 45, n. 11, p. 1759-1769, 2004. <https://doi.org/10.1016/j.enconman.2003.09.019>.

CHEN, J. L.; LI, G. S. Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China. **International Journal of Climatology**, v. 33, p. 487-498, 2013. <https://doi.org/10.1002/joc.3442>.

DELGADO, R. C.; SANTANA, R. O.; GELSLEICHTER, Y. A.; PEREIRA, M. G. Degradation of South American biomes: What to expect for the future?. **Environmental Impact Assessment Review**, v. 96, p. 106815, 2022. <https://doi.org/10.1016/j.eiar.2022.106815>.

ELAGIB, N. A.; MANSELL, M. G. New approaches for estimating global solar radiation across Sudan. **Energy Conversion & Management**, v. 41, n. 5, p. 419-434, 2000. [https://doi.org/10.1016/S0196-8904\(99\)00123-5](https://doi.org/10.1016/S0196-8904(99)00123-5).

ELAGIB, N. A.; BABIKER, S. F.; ALVI, S. H. New empirical models for global solar radiation over Bahrain. **Energy Conversion and Management**, v. 39, p. 827-835, 1998. [https://doi.org/10.1016/S0196-8904\(97\)00035-6](https://doi.org/10.1016/S0196-8904(97)00035-6).

EL-METWALLY, M. Sunshine and global solar radiation estimation at different sites in Egypt. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 67, n. 14, p. 1331-1342, 2005. <https://doi.org/10.1016/j.jastp.2005.04.004>.

EL-SEBAILI, A. A.; AL-GHAMDI, A. A.; AL-HAZMI, F. S.; FAIDAH, A. S. Estimation of global solar radiation on horizontal surfaces in Jeddah, Saudi Arabia.

Energy Policy, v. 37, n. 9, p. 3645-3649, 2009. <https://doi.org/10.1016/j.enpol.2009.04.038>.

ERTEKIN, C. YALDIZ, O. Estimation of monthly average daily global radiation on horizontal surface for Antalya (Turkey). **Renewable Energy**, v. 17, n. 1, p. 95-102, 1999. [https://doi.org/10.1016/S0960-1481\(98\)00109-8](https://doi.org/10.1016/S0960-1481(98)00109-8).

ESCOBEDO, J. F.; GOMES, E. N.; OLIVEIRA, A. P.; SOARES, J. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. **Applied Energy**, v. 86, n. 3, p. 299-309, 2009. <https://doi.org/10.1016/j.apenergy.2008.04.013>.

FALAYI, E. O.; ADEPITAN, J. O.; RABIU, A. B. Empirical models for the correlation of global solar radiation with meteorological data for Iseyin, Nigeria. **International Journal of Physical Sciences**, v.3, n. 9, p. 210-216, 2008. <https://doi.org/10.5897/IJPS.9000218>.

FAN, J.; CHEN, B.; WU, L.; ZHANG, F.; LU, X.; XIANG, Y. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. **Energy**, v. 144, p. 903-914, 2018. <https://doi.org/10.1016/j.energy.2017.12.091>.

GLOVER, J.; MCCULLOCH, J. S. G. The empirical relation between solar radiation and hours of sunshine. **Quarterly Journal of the Royal Meteorological Society**, v. 84, n. 360, p. 172-175, 1958. <https://doi.org/10.1002/qj.49708436011>.

GOODIN, D. G.; HUTCHIMSON, J. M. S.; VANDERLIP, R. L.; KNAPP, M. C. Estimating solar irradiance for crop modeling using daily air temperature data. **Agronomy Journal**, v. 91, n. 5, p. 845-851, 1999. <https://doi.org/10.2134/agronj1999.915845x>.

HARGREAVES, G. L.; ASCE, A. M.; HARGREAVES, G. H.; ASCE, F.; ASCE, RILEY, J. P. Irrigation water requirements for Senegal river basin. **Journal of Irrigation and Drainage Engineering**, v. 111, n. 3, p. 265-275, 1985. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1985\)111:3\(265\)](https://doi.org/10.1061/(ASCE)0733-9437(1985)111:3(265)).

HARGREAVES, G. H.; SAMANI, Z. A. Estimating potential evapotranspiration. **Journal of Irrigation and Drainage Engineering**, v. 108, n. 3, p. 225-230, 1982. <https://doi.org/10.1061/JRCEA4.0001390>.

HASSAN, G. E.; YOUSSEF, M. E.; MOHAMED, Z. E.; ALI, M. A. HANAFY, A. A. New Temperature-based Models for Predicting Global Radiation. **Applied Energy**, v. 179, p. 437-450, 2016. <https://doi.org/10.1016/j.apenergy.2016.07.006>.

INMET, Instituto Nacional de Meteorologia. Normais Climatológicas do Brasil 1991 – 2020. Disponível em: <<https://portal.inmet.gov.br/>>, Acesso em 21 de julho de 2022.

JAHANI, B.; DINPASHOH, Y.; NAFCHI, A. R. Evaluation and development of empirical models for estimating daily solar radiation. **Renewable and Sustainable Energy Reviews**, v. 73, p. 878-891, 2017. <https://doi.org/10.1016/j.rser.2017.01.124>.

KOLEBAJE, O. T.; IKUSIKA, A.; AKINYEMI, P. Estimating solar radiation in Ikeja and Port Harcourt via correlation with relative humidity and temperature. **International Journal of Energy Production and Management**, v. 1, n. 3, p. 253-262, 2016. <https://doi.org/10.2495/EQ-V1-N3-253-262>.

KORACHAGAON, I.; BAPAT, V. N. General formula for the estimation of global solar radiation on earth's surface around the globe. **Renewable Energy**, v. 41, p. 394-400, 2012. <https://doi.org/10.1016/j.renene.2011.11.002>.

LEE, K. H. Improving the correlation between incoming solar radiation and Sunshine hour using DTR. **International Journal of Climatology**, v. 35, n. 3, p. 361-374, 2015. DOI: 10.1002/joc.3983.

LI, M-F.; FAN, L.; LIU, H-B.; GUO, P-T.; WU, W. A general model for estimation of daily global solar radiation using air temperatures and site geographic parameters in Southwest China. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 92, p. 145-150, 2013 (A). <https://doi.org/10.1016/j.jastp.2012.11.001>.

LI, M-F.; TANG, X-P.; WU, W.; LIU, H-B. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. **Energy Conversion and Management**, v. 70, p. 139-148, 2013 (B). <https://doi.org/10.1016/j.enconman.2013.03.004>.

LI, H.; CAO, F.; WANG, X.; MA, W. A temperature-Based model for estimating monthly average daily global solar radiation in China. **The Scientific World Journal**, v. 2014, n. 1, p. 128754, 2014. <https://doi.org/10.1155/2014/128754>.

LI, H.; CAO, F.; BU, X.; ZHAO, L. Models for calculating daily global solar radiation from air temperature in humid regions – A case study. **Environmental Progress & Sustainable Energy**, v. 34, n. 2, p. 595-599, 2015. <https://doi.org/10.1002/ep.12018>.

MARTIM, C. C.; SOUZA, A. P. Estimativas da radiação global com base na insolação na Amazônia brasileira. **Revista Ibero-Americana de Ciências Ambientais**, v. 12, n. 10, 2021. <https://doi.org/10.6008/CBPC2179-6858.2021.010.0020>.

NEWLAND, F. J. A study of solar radiation models for the coastal region of south China. **Solar Energy**, v. 43, n. 4, p. 227-235, 1989. [https://doi.org/10.1016/0038-092X\(89\)90022-4](https://doi.org/10.1016/0038-092X(89)90022-4).

ODODO, J. C.; SULAIMAN, A. T.; AIDAN, J.; YUGUDA, M. M.; OGBU, F. A. The importance of maximum air temperature in the parameterization of solar radiation in Nigeria. **Renewable Energy**, v. 6, n. 7, p. 751-763, 1995. [https://doi.org/10.1016/0960-1481\(94\)00097-P](https://doi.org/10.1016/0960-1481(94)00097-P).

ÖGELMAN, H.; ECEVIT, A.; TASDEMIROGLU, E. A new method for estimating solar radiation from bright sunshine data. **Solar Energy**, v. 33, n. 6, p. 619-625, 1984. [https://doi.org/10.1016/0038-092X\(84\)90018-5](https://doi.org/10.1016/0038-092X(84)90018-5).

PANDAY, C. K.; KATIYAR, A. K. Temperature base correlation for the estimation of global solar radiation on horizontal surface. **International Journal of Energy and Environment**, v. 1, n. 4, p. 737-744, 2010.

PRIETO, J-I.; GARCÍA, D. Global solar radiation models: A critical review from the point of view of homogeneity and case study. **Renewable and Sustainable Energy Reviews**, v. 155, p. 111856, 2022. <https://doi.org/10.1016/j.rser.2021.111856>.

PRESCOTT, J. Evaporation from a Water Surface in Relation to Solar Radiation. **Transactions of the Royal Society of South Australia**, v. 46, p. 114-118, 1940.

QIU, R.; LI, L.; WU, L.; AGATHOKLEOUS, E.; LIU, C.; ZHANG, B.; LUO, Y.; SUN, S. Modeling daily global solar radiation using only temperature data: Past, development, and future. **Renewable and Sustainable Energy Reviews**, v. 163, p. 112511, 2022. <https://doi.org/10.1016/j.rser.2022.112511>.

RAMOS, J. P. A.; VIANNA, M. S.; MARIN, F. R. Estimativa da radiação solar global baseada na amplitude térmica para o Brasil. **Agrometeoros**, v. 26, n. 1, 2018. <http://dx.doi.org/10.31062/agrom.v26i1.26299>.

SABINO, M.; SILVA, A.C.; ALMEIDA, F.T.; SOUZA, A.P. Reference evapotranspiration in climate change scenarios in Mato Grosso, Brazil. **Hydrology**, v. 11, n. 7, 2024. <https://doi.org/10.3390/hydrology11070091>.

SAFFARIPOUR, M. H.; MEHRABIAN, M. A.; BAZARGAN, H. Predicting solar radiation fluxes for solar energy system applications. **International Journal of Environmental Science and Technology**, v. 10, p. 761-768, 2013. <https://doi.org/10.1007/s13762-013-0179-2>.

SAMANTA, S.; BANERJEE, S.; PATRA, P. K.; SEHGAL, V. K.; CHOWDHURY, A.; KUMAR, B.; MUKHERJEE, A. Projection of future daily global horizontal irradiance under four RCP scenarios: An assessment through newly developed temperature and rainfall-based empirical model. **Solar Energy**, v. 227, p. 23-43, 2021. <https://doi.org/10.1016/j.solener.2021.08.049>.

SANTOS, C. M.; TEREMOTO, E. T.; SOUZA, A.; ARISTONE, F.; ILHADDADENE, R. Several models to estimate daily global solar irradiation: adjustment and evaluation. **Arabian Journal of Geosciences**, v. 14, n. 4, p. 286, 2021. <https://doi.org/10.1007/s12517-021-06603-8>.

SOUZA, A. P.; ZAMADEI, T.; MONTEIRO, E. B.; CASAVECCHIA, B. H. Transmissividade Atmosférica da Radiação Global na Região Amazônica de Mato Grosso. **Revista Brasileira de Meteorologia**, v. 31, n. 4, 2016. DOI: <http://dx.doi.org/10.1590/0102-7786312314b20150147>.

SOUZA, A. P.; SILVA, A. C.; TANAKA, A. A.; ULIANA, E. M.; ALMEIDA, F. T.; KLAR, A. E.; GOMES, A. W. A. Global radiation by simplified models for state of Mato Grosso, Brazil. **Pesquisa Agropecuária brasileira**, v. 52, p. 215-227, 2017. DOI: 10.1590/S0100-204X2017000400001.

SOUZA, A. P.; ZAMADEI, T.; BORELLA, D. R. MARTIM, C. C.; ALMEIDA, F. T. ESCOBEDO, J. F. Diurnal Evolution and Estimates of Hourly Diffuse Radiation Based on Horizontal Global Radiation, in Cerrado-Amazon Transition, Brazil. **Atmosphere**, v. 14, n. 8, 2023. <https://doi.org/10.3390/atmos14081289>.

SILVA JUNIOR, C.; LIMA, M. TEODORO, P. E.; OLIVEIRA-JÚNIOR, J. F.; ROSSI, F. S.; FUNATSU, B. M.; BUTTURI, W.; LOURENÇONI, T.; KRAESKI, A.; PELISSARI, T. D.; MORATELLI, F. A.; ARVOR, D.; LUIZ, I. M. S.; TEODORO, L. P. R.; DUBREUIL, V.; TEIXEIRA, V. M. Fires Drive Long-Term Environmental Degradation in the Amazon Basin. **Remote Sensing**, v. 14, n. 2, p. 338, 2022. <https://doi.org/10.3390/rs14020338>.

SWARTMAN, R. K.; OGUNLADE, O. Solar radiation estimates from common parameters. **Solar Energy**, v. 11, n. 3-4, p. 170-172, 1967. [https://doi.org/10.1016/0038-092X\(67\)90026-6](https://doi.org/10.1016/0038-092X(67)90026-6).

TEKE, A.; YILDIRIM, H. B.; ÇELİK, Ö. Evaluation and performance comparison of different models for the estimation of solar radiation. **Renewable and Sustainable Energy Reviews**, v. 50, p. 1097-1107, 2015. <https://doi.org/10.1016/j.rser.2015.05.049>.

THORNTON, P. E.; RUNNING, S. W. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. **Agricultural and Forest Meteorology**, v. 93, n. 4, p. 211-228, 1999. [https://doi.org/10.1016/S0168-1923\(98\)00126-9](https://doi.org/10.1016/S0168-1923(98)00126-9).

TOGRUL, I. T.; ONAT, E. A study for estimating solar radiation in Elazig using geographical and meteorological data. **Energy Conversion & Management**, v. 40, n. 14, p. 1577-1584, 1999. [https://doi.org/10.1016/S0196-8904\(99\)00035-7](https://doi.org/10.1016/S0196-8904(99)00035-7).

TOGRUL, I. T.; TOGRUL, H. EVIN, D. Estimation of global solar radiation under clear sky radiation in Turkey. **Renewable Energy**, v. 21, n. 2, p. 271-287, 2000. [https://doi.org/10.1016/S0960-1481\(99\)00128-7](https://doi.org/10.1016/S0960-1481(99)00128-7).

WEISS, A.; HAYS, C. J.; HU, Q. EASTERLING, W. E. Incorporating bias error in calculating solar irradiance: Implications for crop yield simulations. **Agronomy Journal**, v. 93, n. 6, p. 1321-1326, 2001. <https://doi.org/10.2134/agronj2001.1321>.

WMO. Guide to Meteorological Instruments and Methods of Observation. World Meteorological Organization, 2014.

YILDIRIM, H. B.; TEKE, A.; ANTONANZAS-TORRES, F. Evaluation of classical parametric models for estimating solar radiation in the Eastern Mediterranean region of Turkey. **Renewable and Sustainable Energy Reviews**, v. 82, p. 2053-2065, 2018. <https://doi.org/10.1016/j.rser.2017.08.033>.

ZHANG, J.; ZHAO, L.; DENG, S.; XU, W.; ZHANG, Y. A critical review of the models used to estimate solar radiation. **Renewable and Sustainable Energy Reviews**, v. 70, p. 314-329, 2017. <https://doi.org/10.1016/j.rser.2016.11.124>.

3.9 ANEXO

	MOD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
GRUPO I	1																					
	2																					
	3		■																			
	4		■																			
	5																					
	6		■																			
	7																					
	8																					
	9																					
	10																					
	11																					
	12																					
	13																					
GRUPO II	14																					
	15																					
	16																					
	17																					
	18																					
	19																					
	20																					
	21																					
	22																					
	23																					
	24																					
	25																					
	26																					
	27																					
	28																					
	29																					
	30																					
	31																					
	32																					
	33																					
	34																					
35																						
36																						
37																						
38																						
39																						
40																						
41																						
42																						
43																						
44																						
45																						
GRUPO III	46																					
	47																					
	48																					
	49																					
GRUPO IV	50																					
	51																					
	52																					
	53																					
	54																					
	55																					
GRUPO V	56																					
	57																					
	58																					
	59																					
	60																					
	61																					
	62																					
	63																					
	64																					
	65																					
	66																					
	67																					
	68																					
69																						
70																						
71																						
72																						
73																						
74																						
75																						
76																						
77																						
78																						
79																						
80																						
81																						
82																						
83																						
84																						
85																						
86																						
87																						

Figura 3.9. Teste-t dos coeficientes dos modelos, com 0,05 de significância. (não significativo - NS ■).

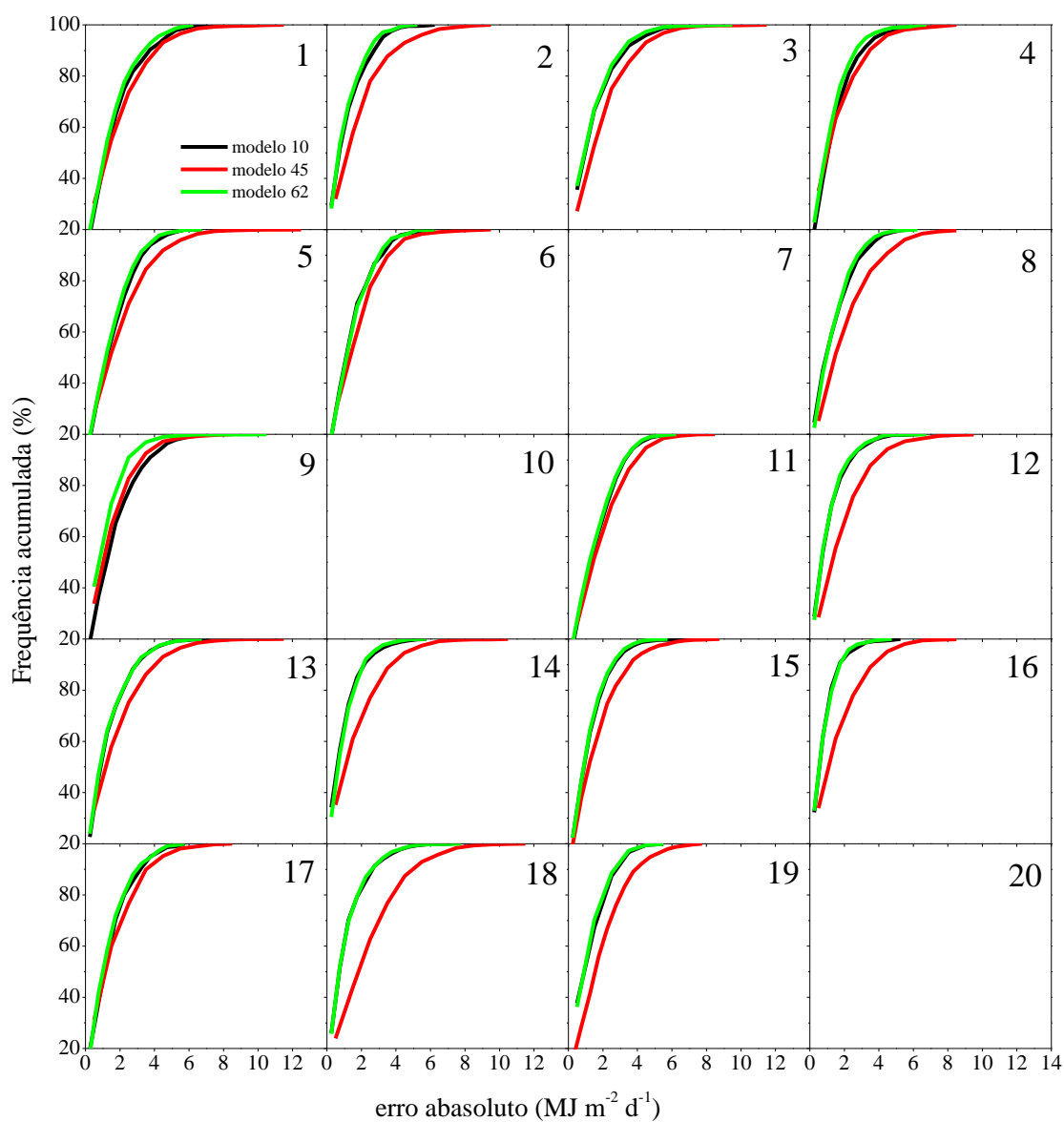


Figura 3.10. Frequência acumulada do erro absoluto dos modelos 10, 45 e 62 nas 20 estações meteorológicas localizadas no Bioma Amazônico brasileiro.

4.0 CAPÍTULO 2 - MACHINE LEARNING NA ESTIMATIVA DA RADIAÇÃO GLOBAL DIÁRIA NA AMAZÔNIA BRASILEIRA

RESUMO: O conhecimento da radiação global (Hg) é primordial para o desenvolvimento econômico regional, e pode auxiliar no direcionamento de políticas públicas relacionadas ao potencial agropecuário e energético. No entanto, ainda existem limitações na sua disponibilidade em diversas regiões brasileiras. Este trabalho avalia a capacidade preditiva de duas técnicas de Machine Learning (ML) como Multi Layer Perceptron (MLP) e Máquina de Vetor e Suporte (SVM) na estimativa de Hg em 20 estações meteorológicas com 40 diferentes combinações de entrada envolvendo insolação, temperatura do ar, umidade relativa do ar, fotoperíodo e radiação extraterrestre, e também é comparado com três modelos empíricos baseados na insolação, temperatura e combinação híbrida. No geral, quanto maior o número de variáveis de entrada, melhor o desempenho das técnicas de ML, principalmente em combinações envolvendo a insolação que reduziu a dispersão da Hg estimada nos dias com alta transmissividade atmosférica e a temperatura do ar em dias com baixa transmissividade atmosférica. O desempenho do SVM foi melhor quando comparado ao MLP em todos os indicadores estatísticos. As técnicas de ML e os modelos empíricos apresentaram desempenho semelhantes quando empregado a insolação e temperatura do ar como variável de entrada na estimativa da radiação global diária. Sendo assim, em função da fácil implementação, é recomendado o emprego dos modelos empíricos na estimativa da radiação global diária na Amazônia brasileira quando tem disponibilidade de dados da insolação e temperatura do ar.

Palavras-chave: SVM, MLP, inteligência artificial, energia solar, indicadores estatísticos, transmissividade atmosférica.

4.1 INTRODUÇÃO

No contexto mundial, a demanda por fontes de energia renováveis com baixa emissão de carbono vem crescendo, sendo que vários países estão aproveitando o potencial solar para a implementação de projetos fotovoltaicos, porém é necessária avaliação da viabilidade de cada região. Nesse contexto, o Brasil é um país que

apresenta grande potencial de aproveitamento da energia solar, para projetos fotovoltaicos, pois sua localização geográfica privilegiada garante abundante incidência de radiação global (Hg), em parte considerável do seu território, e ao longo do ano. Atualmente, a energia solar representa em torno de 13% de toda a matriz elétrica brasileira, sendo a segunda maior fonte do país, ficando atrás apenas da energia hídrica (ABSOLAR, 2024).

Existem diversos benefícios econômicos e ambientais que ajudam a impulsionar o crescimento dessa fonte de energia renovável no Brasil. Majoritariamente, a energia solar está sendo utilizada no setor residencial residências como alternativa, que pode auxiliar na redução dos custos com energia elétrica, seja por meio da energia térmica (aquecimento de água) ou com a utilização de energia fotovoltaica (gerando eletricidade). Nos últimos anos, principalmente na região Centro-Oeste e Sul da Amazônia brasileira, com a consolidação e avanço das áreas de produção agropecuária, têm-se estabelecido projetos de plantas de geração fotovoltaica em propriedades agrícolas e agroindustriais, objetivando o abastecimento de energia em sistemas de irrigação, armazéns e secadores, instalações zootécnicas, sedes e residências, dentre outras infraestruturas rurais que demandam eletricidade.

O conhecimento da radiação global incidente na superfície é fundamental, pois é uma informação estratégica e necessária para o planejamento de várias atividades, como os sistemas agrícolas, determinação da evapotranspiração potencial, modelagem do crescimento das culturas, dimensionamento de sistemas energéticos, monitoramento das mudanças climáticas, ecologia, construção entre outros (ALLEN et al. 1998; FAN et al. 2019; HE et al. 2020; AGBULUT et al. 2021; BELLIDO-JIMÉNEZ et al. 2021; BOUNOUA et al. 2021; GÜREL et al. 2023).

No contexto nacional, devido à grande extensão territorial do Brasil e dificuldades de acesso, logística e de recursos financeiros e humanos para instalação e manutenção de sensores de medição, na grande maioria das estações meteorológicas brasileiras, a radiação global é a variável meteorológica com menor disponibilidade de dados contínuos e consistentes. Essa realidade também é observada em outras regiões do mundo, visto que os piranômetros, dependendo do modelo e do objetivo de monitoramento, apresentam altos custos de aquisição e necessitam de manutenções periódicas (BOUNOUA et al. 2021; ZHOU et al. 2021).

O bioma Amazônia abrange aproximadamente 50% do território brasileiro e possui uma área de 4.196.943 km²; é considerada como a maior floresta tropical do mundo, e possui grande estoque de carbono na vegetação e no solo, além de assimilar continuamente o CO₂ da atmosfera por meio da realização da fotossíntese da vegetação. Nessa região, o monitoramento meteorológico de superfície, é realizado com 72 estações meteorológicas automáticas (EMAs), e 20 estações meteorológicas convencionais (EMCs) pertencentes à Rede de Estações do Instituto Nacional de Meteorologia (INMET); além dessas estações, ainda existem medidas com séries temporais menores em universidades e instituições de pesquisas públicas e privadas.

Em função da dificuldade na mensuração da radiação global devido principalmente aos custos envolvidos na aquisição dos sensores (NAWAB et al. 2023), várias pesquisas avaliaram, ao longo dos anos, diferentes metodologias de estimativas de Hg, através da análise de correlação com variáveis meteorológicas de obtenção menos limitadas, como por exemplo a insolação, temperatura do ar, umidade relativa do ar, dentre outros (AGBULUT et al. 2021; BOUNOUA et al. 2021; ZHOU et al. 2021; GÜREL et al., 2023). As metodologias mais difundidas na estimativa da Hg são os modelos empíricos, porém recentemente têm aumentado o número de trabalhos que avaliam as técnicas de Machine Learning e suas respostas nessas estimativas (NAWAB et al. 2023; MARQUES et al. 2024). O primeiro modelo empírico foi proposto no ano de 1924, com base na regressão linear entre a radiação global e a insolação, conhecido como modelo de Angström-PreScott, porém ao longo do tempo surgiram propostas de alterações nesse modelo e a geração de inúmeros outros modelos, com novas funções analíticas e variáveis de entrada (BOUNOUA et al. 2021).

O aumento da capacidade de processamento dos sistemas computacionais propiciou o desenvolvimento de novas metodologias de estimativas, com destaque para as técnicas de Machine Learning (ZHOU et al. 2021). O aprendizado de máquina ou Machine Learning (ML) é muito utilizado em previsões de eventos com características não-lineares, no qual o processo de aprendizagem é iniciado após o fornecimento de uma base de dados para o treinamento e a técnica mapeia os padrões que serão utilizados para prever valores futuros (AGBULUT et al. 2021). Como boa parte dos problemas ambientais possui componentes não lineares entre as variáveis

dependentes e independentes em função dos ruídos, tem aumentado o número de estudos empregando os modelos de ML, com aplicação em várias áreas do conhecimento; dentre elas, está a micrometeorologia, em específico a estimativa da radiação global (HE et al. 2020).

Recentemente, Zhou et al. (2021) revisaram 232 artigos envolvendo essa temática, e observaram um crescimento exponencial da utilização dessas técnicas para estimativas de Hg, entre os anos de 2001 e 2020. Esses autores relataram que existem diferentes tipos de variáveis de entrada, como variáveis meteorológicas, poluição do ar, parâmetros geográficos, parâmetros de calendário e parâmetros astronômicos, como radiação extraterrestre, declinação solar, ângulo zenital e azimute. Segundo Nawab et al. (2023), as variáveis meteorológicas mais utilizadas na estimativa da Hg é a temperatura do ar, umidade relativa do ar, transmissividade atmosférica e precipitação pluvial. Marques et al. (2024) ainda destacam que, em função das variações espaciais e sazonais dos elementos meteorológicos, deve-se avaliar individualmente, para cada região, quais variáveis de entrada e metodologias apresentam melhores respostas nas estimativas de Hg.

As principais técnicas de ML empregados na estimativa da radiação global são: Máquina de Vetores e Suporte (SVM) e o Multi Layer Perceptron (MLP) (QUEJ et al. 2017; SILVA et al. 2017; ANTONOPOULOS et al. 2019; FAN et al. 2019; FENG et al. 2019; AGBULUT et al. 2021; BELLIDO-JIMÉNEZ et al. 2021; HUSAIN & KHAN, 2021; MARQUES et al. 2024). De 2000 a 2014, o MLP era predominante utilizado em todas as publicações, porém a partir de 2009, o SVM começou a ser mais empregado (VOYANT et al. 2017). Ambas as técnicas, são indicados para solução de problemas complexos envolvendo diversas variáveis, e exigem baixo esforço computacional no processamento. As técnicas foram avaliados em vários países, como no Brasil (SILVA et al. 2017; SANTOS et al. 2021), Turquia (KABA et al. 2018; AGBULUT et al. 2021; KÜÇÜKTOPÇU et al. 2024), Espanha (BELLIDO-JIMÉNEZ et al. 2021), EUA (BELLIDO-JIMÉNEZ et al. 2021), Marrocos (BOUNOUA et al. 2021), China (FAN et al. 2019; FENG et al. 2019; HE et al. 2020), Iran (HUANG et al. 2022), Índia (HUSAIN & KHAN, 2021), México (QUEJ et al. 2017), Marrocos (MARZOUQ et al. 2019), Grécia (ANTONOPOULOS et al. 2019) e Etiópia (WOLDEGIYORGIS et al. 2023). No bioma Amazônico, recentemente Marques et al.

(2024) avaliaram as estimativas de Hg para 12 locais inseridos no estado do Amazonas, com uma única combinação de variáveis de entrada, e por fim, esses autores recomendaram que em trabalhos futuros deveria ser avaliada a performance de diferentes combinações de entrada.

Nesse contexto, objetivou-se avaliar a influência de diferentes combinações de entrada na estimativa da radiação global empregando as técnicas de MLP e SVM e também comparando com modelos empíricos, no bioma Amazônico brasileiro.

4.2 MATERIAL E MÉTODOS

4.2.1 Área de estudo

O bioma Amazônico brasileiro possui área de aproximadamente 4.196.943 km² abrangendo os territórios dos estados do Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e parcialmente os estados de Tocantins, Mato Grosso e Maranhão. O monitoramento meteorológico na região é efetuado com aproximadamente 72 estações meteorológicas automáticas (EMAs) e 20 estações meteorológicas convencionais (EMCs) sob a responsabilidade do Instituto Nacional de Meteorologia (INMET), os dados estão disponíveis on-line e podem ser baixados no endereço eletrônico (<https://portal.inmet.gov.br/>). Neste estudo, foram utilizados dados de um total de 20 estações meteorológicas convencionais (EMCs) e automáticas (EMAs), distribuídas no bioma Amazônico brasileiro (Figura 4.1).

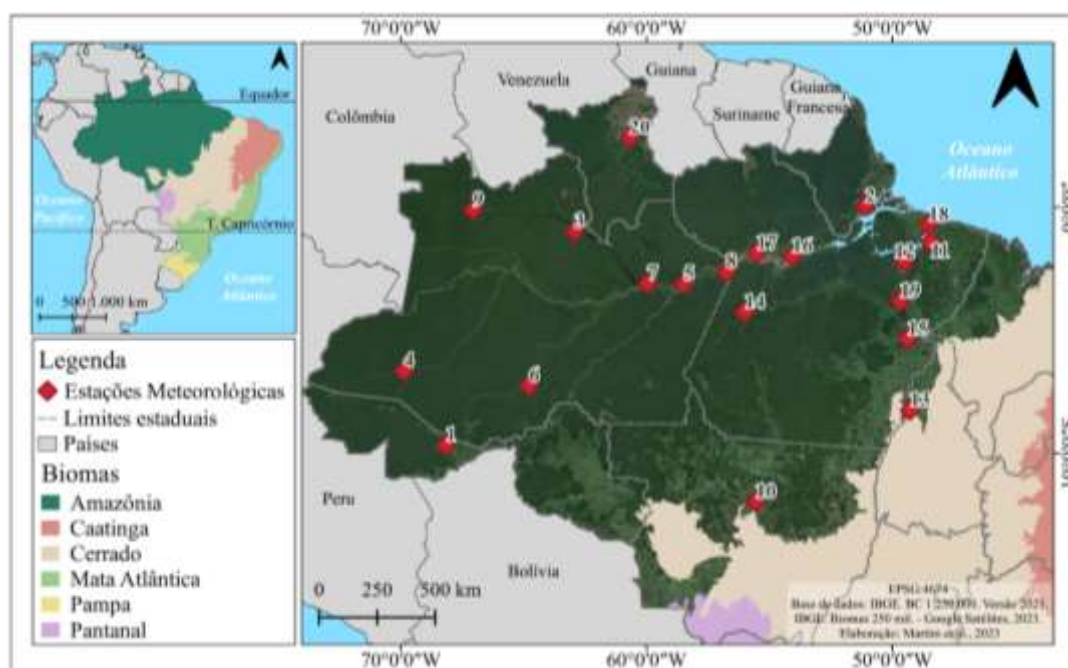


Figura 4.1. Mapa de localização das 20 estações meteorológicas automáticas e convencionais distribuídas pelo bioma Amazônia.

As informações gerais das estações meteorológicas como localização geográfica, classificação climática onde estão inseridas e período de funcionamento são apresentadas na Tabela 4.1. A amplitude da série histórica de dados varia de 5 anos para estação da cidade de Óbidos até 22 anos para cidade de Manaus, e de acordo com a classificação de Köppen as estações estão inseridas em três climas, sendo o clima tropical de monção (Am) com temperatura média mensal acima de 18 °C, precipitação pluvial média anual é maior que 1500 mm e o mês mais seco é menor que 60 mm, clima tropical úmido (Af) com temperatura do ar média mensal acima de 18 °C, precipitação pluvial média mensal acima de 60 mm e, clima tropical de savana (Aw) com temperatura média mensal acima de 18 °C com chuvas no verão (ALVARES et al. 2013).

Tabela 4.1. Estações meteorológicas automáticas inseridas na Amazônia brasileira.

Estado	Cidade	CCKP*	Lat.	Lon.	Alt.	Período de funcionamento
Acre	1 - Rio Branco	Am	-9,67	-68,16	163	2015-2022
Amapá	2 - Macapá	Am	0,035	-51,08	16	2013-2022
Amazonas	3 - Barcelos	Af	-0,98	-62,92	29	2008-2022
	4 - Eirunepé	Af	-6,65	-69,87	121	2012-2022
	5 - Itacoatiara	Af	-3,12	-58,47	41	2008-2022
	6 - Lábrea	Am	-7,25	-64,78	61	2008-2018

	7 – Manaus	Af	-3,1	-59,95	61	2000-2022
	8 – Parintins	Af	-2,63	-56,75	18	2008-2018
	9 - São Gabriel da Cachoeira	Af	-0,12	-67,05	79	2011-2022
Mato Grosso	10 – Sinop	Aw	-11,97	-55,55	366	2006-2017
	11 - Belém	Af	-1,41	-48,43	21	2003-2022
	12 - Cametá	Af	-2,23	-49,48	9	2008-2022
	13 - Conceição do Araguaia	Aw	-8,25	-49,27	175	2008-2022
	14 - Itaituba	Af	-4,27	-56,00	24	2008-2022
Pará	15 - Marabá	Aw	-5,36	-49,37	116	2009-2022
	16 - Monte Alegre	Am	-2,0	-54,07	100	2012-2022
	17 - Óbidos	Am	-1,88	-55,51	89	2012-2017
	18 - Soure	Am	-0,72	-48,51	12	2008-2017
	19 - Tucuruí	Am	-3,82	-49,67	137	2008-2017
Roraima	20 - Boa Vista	Am	2,82	-60,68	82	2010-2022

Latitude (Lat.), Longitude (Lon.), Altitude (Alt.), *Alvares et al. (2013)

4.2.2 Análise dos dados

As variáveis meteorológicas diárias selecionadas para este estudo foram a temperatura máxima (Tmax), temperatura média (Tmed), temperatura mínima (Tmin), umidade relativa máxima (RHmax), umidade relativa média (RHmed), umidade relativa mínima (RHmin) e radiação global (Hg) obtidas nas EMAs e a insolação (S) obtida nas EMCs. Para padronizar os dados de entrada, todas as variáveis empregadas sendo meteorológicas ou astronômicas, foram integradas na escala de tempo diária.

Além das variáveis medidas através das EMAs, duas várias astronômicas também foram utilizadas nas diferentes combinações, a radiação solar extraterrestre (Ho) e o fotoperíodo (So), sendo que essas variáveis são dependentes da época do ano e da latitude, podendo ser obtidas por meio das equações abaixo (ALLEN et al. 1998).

$$Ho = 37,59 * dr * \left(\frac{\pi}{180} * \sigma * \sin \phi * \sin \delta + \cos \phi * \cos \delta * \sin \sigma \right) \quad 4.1$$

$$dr = 1 + 0,033 * \cos \left(\frac{360 * DJ}{365} \right) \quad 4.2$$

$$\delta = 23,45 * \sin \left[\frac{360}{365} (DJ + 284) \right] \quad 4.3$$

$$\sigma = \cos^{-1}(-\tan \phi * \tan \delta) \quad 4.4$$

$$So = \frac{2 * \sigma}{15} \quad 4.5$$

em que: 37,6 (constante); ϕ é a latitude local (°); δ é declinação solar (°, Equação 4.3); σ é o ângulo horário diário (°, Equação 4.4); dr é o fator de excentricidade da órbita da

terra (Equação 4.2); e dia juliano (DJ) que é ordenamento dos dias ao longo do ano (dias) começando de 1 em 1º de janeiro a 365 em 31 de dezembro ($1 \leq DJ \leq 365$). E a S_o pode ser obtida por meio da equação 4.5.

Durante o processo de treinamento e validação as técnicas de ML, além da disponibilidade das séries históricas, também existe a necessidade de avaliar a qualidade dos dados; nesse caso, deve ser levado em consideração a implementação de filtros rigorosos para evitar valores com falhas de leitura ou inconsistentes (Zhou et al., 2021). Sendo assim, os dados foram submetidos a filtros, sendo excluídos todos os dados do mesmo dia, caso alguma das seguintes condições fossem atendidas: i) a transmissividade atmosférica ($K_t = H_g/H_o$) acima de 0,85; ii) razão de insolação (S/S_o) maior que 1; iii) falha da H_g horária entre 9:00 e 15:00 (horário local); iv) falha no valor diário de T_{max} , T_{med} , T_{min} , RH_{max} , RH_{med} , RH_{min} e S .

Devido a divergência entre as unidades que representam as variáveis de entrada, elas foram submetidas a um processo de normalização para que os valores de saída ficassem normalmente distribuídos ao longo de suas variações e entre -1 a 1, e adimensionais em cada variável. De acordo com Bellido-Jiménez et al. (2021) e He et al. (2020), este é um procedimento comum em trabalhos que envolvem ML. Na avaliação dos modelos de ML foi empregado o aprendizado do tipo supervisionado, com separação das bases de dados de cada estação em 70% para treinamento e 30% para teste (avaliação do desempenho estatístico), sistematicamente ao longo da série histórica disponível, para garantir a representatividade e proporcionalidade dos períodos (semanas, meses e anos).

4.2.3 Inteligência artificial (IA) e Machine Learning (ML)

A inteligência artificial (IA) está presente em diversos processos, com o objetivo de resolver problemas complexos, pois permite que o sistema possa tomar decisões de forma autônoma, de acordo com um aprendizado predefinido (NAWAB et al. 2023). A IA é uma grande área do conhecimento que originou o que conhecemos atualmente como machine learning (ML), sendo dividido em várias técnicas, que possui capacidade de encontrar padrões e detectar tendências em sistemas não lineares, em problemas que envolvem classificação e regressão (AGBULUT et al. 2021; FAN et al. 2019). As técnicas mais populares são a Rede Neural Artificial (ANN) e Máquina de Vetores de Suporte (SVM), sendo que essas técnicas permitem o processamento de

problemas complexos que envolvem dezenas, centenas ou milhares de variáveis (Big Data), como é o comportamento da maioria dos problemas da atualidade.

4.2.3.1 Rede Neural Artificial (ANN) do tipo Multi Layer Perceptron (MLP)

A Rede Neural Artificial (ANN) foi desenvolvida como base em observações de estudos envolvendo modelos de neurônios do sistema nervoso biológico e forneceu toda a base teórica para o que conhecemos hoje como rede neural artificial (AGBULUT et al. 2021; BELLIDO-JIMÉNEZ et al. 2021; BOUNOUA et al. 2021; FAN et al. 2019).

A ANN mais utilizada é o Multi Layer Perceptron (MLP), principalmente na resolução de problemas complexos com inúmeras variáveis, que requer um baixo poder computacional, viabilizando a modelagem e análise de bancos de dados extensos (FAN et al. 2019). De forma básica a estrutura do MLP é subdividida em três camadas: i) camada de entrada (CE), que no caso deste estudo são as variáveis meteorológicas (X); ii) camada oculta (CO), onde se encontram os neurônios (n), sendo que a quantidade de “n” depende da complexidade do problema e do número das variáveis de entrada; iii) camada de saída (CS), que representa o resultado final do MLP, que neste caso, é a estimativa de Hg (MARQUES et al. 2024). Todas as camadas são interligadas por pesos (W), que são valores numéricos que inferem a importância de cada variável de entrada na variável de saída, gerando assim conexões ponderadas (BELLIDO-JIMÉNEZ et al. 2021; FAN et al. 2019). Matematicamente, o valor de saída da MLP pode ser modelado pela Equação 4.6, com a multiplicação dos W por cada variável X e o viés (β), onde a somatória é submetida a função de ativação sigmoide (Equação 4.7) não linear com valores de saída oscilando de 0 a 1 para cada neurônio e o k é o resultado em cada camada oculta (Equação 4.8) respectivamente.

$$y_X = f\left(\sum_{j=1}^n W_{i,n} X_{i,n} + \beta_i\right) \quad 4.6$$

$$f(k) = \frac{1}{(1+e^{-k})} \quad 4.7$$

$$k = \sum_{j=1}^n W_{i,n} X_{i,n} + \beta_i \quad 4.8$$

em que: y_X é a variável de saída; X é a variável de entrada; $f(k)$ é a função de transformação sigmoide em cada camada oculta; W são os pesos; k é o resultado em cada camada oculta (Equação 4.8); β_i é a bias/viés em cada camada.

No processo de treinamento para o ajuste de “W”, o algoritmo iterativo Back Propagation (BP) tem por objetivo a minimização da função de perda usando gradiente descendente para aprendizagem supervisionada; esse processo de treinamento inicia de acordo com a Equação 4.6, seguindo o sentido direto conhecido como Feed-Forward, já no sentido inverso, a diferença entre o valor esperado (MED) e o resultado de saída (EST), determina o erro (E) da estimativa representado na Equação 4.9, que serve como referência para a atualização de “W”, processo esse conhecido como Feed-Backward; esse processo se repetirá iterativamente até encontrar o erro mínimo pela redução da função de perda ou até que atinja o número de interações pré-determinadas para que o aprendizado do MLP aconteça. Após o treinamento, o MLP será capaz de forma rápida e confiável prever valores de Hg, mesmo que o conjunto de dados de entrada contenha ruído (AGBULUT et al. 2021).

$$E = \sum_{i=1}^n \frac{(\text{MED}-\text{EST})^2}{2} \quad 4.9$$

Uma das etapas de maior relevância quando se trabalha com técnicas de ML é a definição dos melhores hiperparâmetros, como a arquitetura e as configurações que devem ser determinadas antes do treinamento (BELLIDO-JIMÉNEZ et al. 2021). No caso do MLP, é o número de variáveis de entrada, neurônios, camada oculta, função de ativação, algoritmo de otimização para o treinamento entre outros. Neste trabalho, através de pré-testes optou-se por utilizar apenas uma camada oculta (Figura 4.2), com o número de neurônios (n) variando de acordo com a Equação 10, dada pela soma entre VE (número de variáveis de entrada) e VS (número de variáveis de saída). A taxa de aprendizagem, momento e número de interações empregados foram de 0,3; 0,2 e 1000 (SANTOS et al. 2021; SILVA et al. 2017).

$$n = (\text{VE} + \text{VS}) \quad 4.10$$

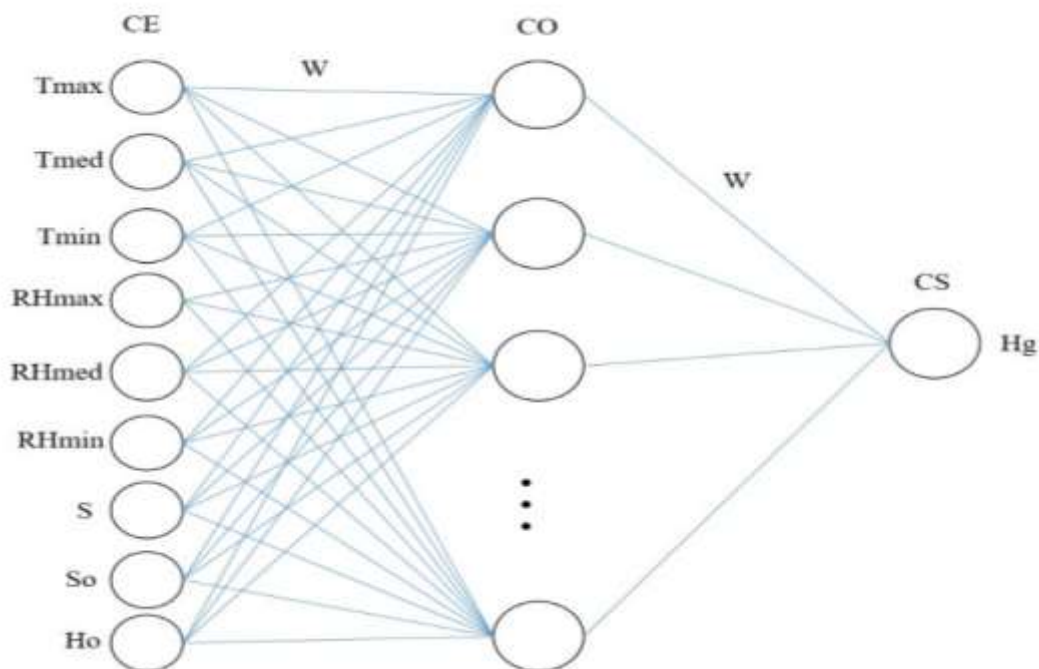


Figura 4.2. Estrutura organizacional da rede neural artificial desenvolvida para MLP.

4.2.3.2 Máquina de Vetores de Suporte (SVM)

A Máquina de Vetores de Suporte ou Support Vector Machine (SVM) é uma técnica de ML supervisionado proposto por Vapnik (1995) com inúmeras aplicações em problemas reais, sendo muito eficiente em problemas envolvendo classificação e regressão (QUEJ et al. 2017).

Como demonstrado na Figura 4.3, no SVM é traçado o hiperplano que melhor separa as diferentes classes, e, os dados que estão próximos do hiperplano, são conhecidos como vetores de suporte; sendo que nesse caso, o hiperplano pode ser uma reta ou plano com a função de separar as diferentes classes (AGBULUT et al. 2021; BELLIDO-JIMÉNEZ et al. 2021). O algoritmo engloba diversas funções de ativação, sendo que a mais utilizada é a função de base radial (RBF) de Kernel, pois ela é de fácil implementação, eficiente e pode ser utilizada em problemas multidimensionais, ou seja, com um número grande de variáveis de entrada (FAN et al. 2019). Na função RBF de Kernel, deve ser fornecido alguns parâmetros que podem sofrer alterações de acordo com a variável de entrada, como o custo (C), épsilon (ϵ) e gama (γ) (QUEJ et al. 2017). Após vários pré-testes foram definidos os seguintes valores dos parâmetros (C = 100, ϵ = 0,001, γ = 0,3), corroborando com Silva et al. (2017) e Santos et al.

(2021), em regiões próximas. Na Equação 4.11, y_r é o valor de saída, ω e b são conhecidos como peso do vetor e do viés e $\text{RBF} = (X_1, X_2)$ como a função não linear.

$$y_r = \omega \text{RBF}(x_1, x_2) + b \quad 4.11$$

Na função RBF de Kernel descrita pela Equação 4.12, o parâmetro $\|X_1 - X_2\|^2$ representa a distância euclidiana elevada ao quadrado no espaço de entrada e $\gamma = -1/2 \sigma^2$ é o valor determinado por σ que é um parâmetro livre, e representa o ruído gaussiano padrão num espaço de dimensão infinita (ZHOU et al. 2021).

$$\text{RBF}(X_1, X_2) = \exp(-\gamma \|X_1 - X_2\|^2) \quad 4.12$$

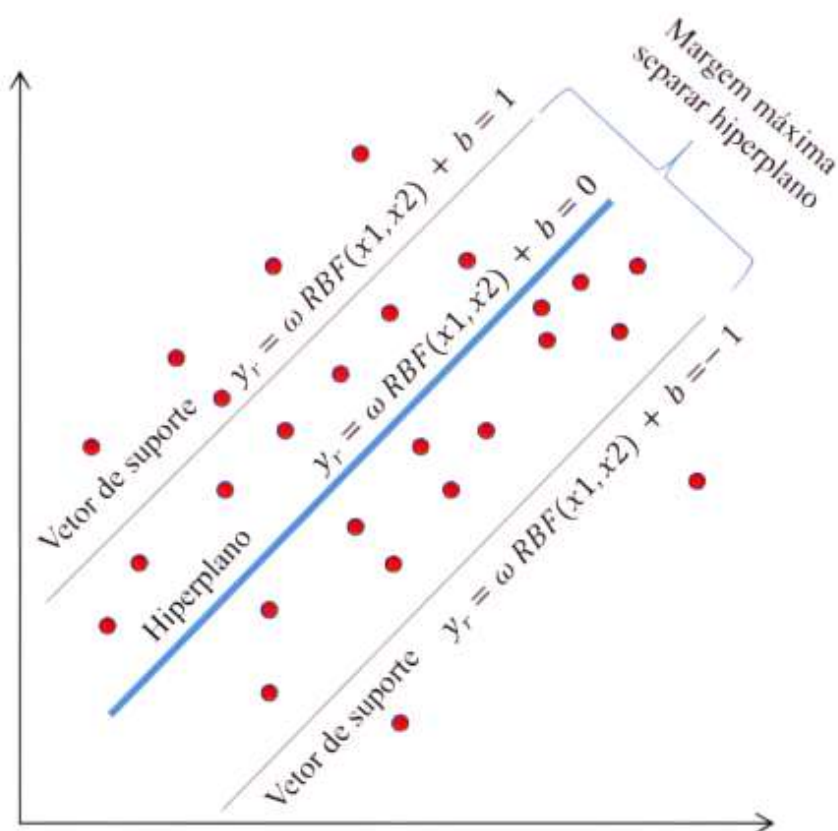


Figura 4.3. Máquina de vetores de suporte (SVM).

No treinamento do SVM, foi empregado o algoritmo de Otimização Sequencial Mínima (SMO), que é de fácil implementação, com baixo esforço computacional e muito utilizado para a solução de problemas que envolvem regressão; e através de iterações obtém a melhor solução, com multiplicadores de Lagrange (SILVA et al. 2017; SHEVADE et al. 2000).

4.2.4 Estrutura dos modelos de ML

A correta seleção das variáveis de entrada representa um ponto importante a ser considerado pois influencia significativamente na capacidade preditiva. Vários estudos avaliaram o impacto das diferentes combinações no desempenho das técnicas de ML, e, em alguns casos a redução do número de variáveis meteorológicas melhorou a capacidade preditiva do modelo de ML (HE et al. 2020; BOUNOUA et al. 2021; KABA et al. 2018).

Os dados de entrada foram divididos em 40 diferentes combinações empregando as técnicas MLP (MLP1 a MLP40) e SVM (SVM1 a SVM40) e em sete tipos de agrupamentos de combinações, de acordo com a variável de entrada: I) S, So e Ho (combinação 1), II) Tmax, Tmed, Tmin, So e Ho (combinação 2 a 7), III) RHmax, RHmed, RHmin, So e Ho (combinação 8 a 10), IV) Tmax, Tmed, Tmin, RHmax, RHmed, RHmin, So e Ho (combinação 11 a 24), V) Tmax, Tmed, Tmin, S, So e Ho (combinação 25 a 29), XI) RHmax, RHmed, RHmin, S, So e Ho (combinação 30 a 34), XIII) Tmax, Tmed, Tmin, RHmax, RHmed, RHmin, S, So e Ho (combinação 35 a 40) (Tabela 4.2).

Tabela 4.2. Diferentes combinações das variáveis de entrada para estimativas de Hg pelas técnicas MLP e SVM, no bioma Amazônico brasileiro.

Nº	Variáveis meteorológicas
1	S, So, Ho
2	Tmax, So, Ho
3	Tmed, So, Ho
4	Tmax, Tmed, So, Ho
5	Tmax, Tmin, So, Ho
6	Tmed, Tmin, So, Ho
7	Tmax, Tmed, Tmin, So, Ho
8	RHmed, So, Ho
9	RHmin, So, Ho
10	RHmax, RHmed, RHmin, So, Ho
11	Tmax, RHmax, So, Ho
12	Tmax, RHmed, So, Ho
13	Tmax, RHmin, So, Ho
14	Tmed, RHmed, So, Ho
15	Tmed, RHmed, So, Ho
16	Tmed, RHmin, So, Ho
17	Tmin, RHmed, So, Ho
18	Tmin, RHmin, So, Ho

19	Tmax, Tmin, RHmax, RHmin, So, Ho
20	Tmax, Tmed, Tmin, RHmed, So, Ho
21	RHmax, RHmed, RHmin, Tmed, So, Ho
22	RHmax, RHmed, RHmin, Tmax, Tmin, So, Ho
23	Tmax, Tmed, Tmin, RHmax, RHmin, So, Ho
24	RHmax, RHmed, RHmin, Tmax, Tmed, Tmin, So, Ho
25	Tmax, S, So, Ho
26	Tmed, S, So, Ho
27	Tmin, S, So, Ho
28	Tmax, Tmin, S, So, Ho
29	Tmax, Tmed, Tmin, S, So, Ho
30	RHmax, S, So, Ho
31	RHmed, S, So, Ho
32	RHmin, S, So, Ho
33	RHmax, RHmin, S, So, Ho
34	RHmax, RHmed, RHmin, S, So, Ho
35	Tmax, Tmin, Rhmax, S, So, Ho
36	Tmax, Tmin, Rhmin, S, So, Ho
37	RHmax, RHmin, Tmax, Tmin, So, Ho
38	RHmax, RHmed, RHmin, Tmax, Tmin, S, So, Ho
39	RHmax, RHmin, Tmax, Tmed, Tmin, S, So, Ho
40	RHmax, RHmed, RHmin, Tmax, Tmed, Tmin, S, So, Ho

Tanto os modelos MLP's como os SVM's foram implementados no software de código aberto Waikato Environment for Knowledge Analysys (WEKA), que possui várias bibliotecas de ML, que preparam os dados, resolvem problemas de regressão, problemas de classificação, visualização, mineração e associação, sendo intuitivo e de fácil execução (<https://www.cs.waikato.ac.nz/ml/index.html>). No WEKA, foi utilizado o pacote SMOreg para treinamento e validação da técnica SVM e o pacote MultilayerPerceptron para a técnica MLP.

4.2.5 Modelos empíricos

Para comparação conjunta dos modelos empíricos com as técnicas de ML, foram selecionados quatro modelos empíricos calibrados localmente (mesmas estações) e recomendados por Martim et al. (2024) para o bioma Amazônico; nesse caso, foram considerados modelos que apresentam como variáveis de entrada a insolação (S) (Equação 4.13), a amplitude térmica (ΔT) e a temperatura média diária

(Tmed) (Equação 14) e um modelo híbrido com amplitude térmica e insolação (Equação 4.15).

$$\frac{Hg}{Ho} = a + b \left(\frac{S}{So} \right)^c \quad 4.13$$

$$\frac{Hg}{Ho} = a + b \Delta T + b \Delta T^{0,25} + d \Delta T^{0,5} + \frac{eTmed}{Ho} \quad 4.14$$

$$\frac{Hg}{Ho} = a + b \ln \Delta T + c \left(\frac{S}{So} \right)^d \quad 4.15$$

4.2.6 INDICATIVOS ESTATÍSTICOS

Na avaliação do desempenho, existem vários indicativos estatísticos, entre os mais utilizados se destaca o erro médio absoluto (MAE - Equação 4.16), raiz quadrada do erro médio (RMSE, Equação 4.17), erro relativo médio (MBE - Equação 4.18), índice de concordância de Willmott (d - Equação 4.19) e coeficiente de determinação (R^2 - Equação 4.20). O coeficiente de correlação de Pearson (r), representado pela Equação 4.21, indica o grau de correlação linear entre duas variáveis (LIU et al. 2023), sendo que os valores variam de -1 a 1, sendo que o zero indica nenhuma correlação, o 1 indica correlação perfeita, e -1 indica correlação inversa perfeita. Quanto maior a correlação, maior será a importância desta variável de entrada na variável de saída (MARQUES et al. 2024).

$$MAE = \left[\frac{\sum_{i=1}^N |p_i - o_i|}{N} \right] \quad 4.16$$

$$RMSE = \left[\frac{\sum_{i=1}^N (P_i - O_i)^2}{N} \right]^{\frac{1}{2}} \quad 4.17$$

$$MBE = \left[\frac{\sum_{i=1}^N (P_i - O_i)}{N} \right] \quad 4.18$$

$$d = 1 - \left[\frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i - O_i| + |O_i - O|)^2} \right] \quad 4.19$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (O_i - O)^2} \quad 4.20$$

$$r = \frac{\sum (O_i - O)(P_i - P)}{\sqrt{\sum (O_i - O)^2 \sum (P_i - P)^2}} \quad 4.21$$

em que: P_i é o valor estimado; P é o valor médio estimado; O_i é o valor de referência das estações meteorológicas; O é a média dos valores de referência; N é o número total de observações.

4.3 RESULTADOS

As correlações entre todas as variáveis meteorológicas medidas nas 20 estações localizadas no bioma Amazônico brasileiro, foram analisadas através do coeficiente de correlação de Pearson (r) (Figura 4.4). Os valores de correlação maiores que 0,5 e menor que -0,5 são classificados como correlação forte; entre 0,3 a 0,5 ou -0,3 a -0,5 apresentam correlação fraca e abaixo de 0,3 ou -0,3 não apresentam correlação (LIU et al., 2023). A correlação entre Hg e a temperatura do ar foi positiva com valor de 0,66; 0,56 e 0,081 para Tmax, Tmed e Tmin, respectivamente; já com a umidade relativa, a correlação foi inversa com valor de -0,58; -0,52 e -0,23 para RHmax, RHmed e RHmin, respectivamente; já para a insolação (S), a correlação chegou a 0,83. No geral, o valor absoluto dessas correlações de Hg com as outras variáveis meteorológicas podem ser classificados em ordem crescente da menor para a maior correlação como $|Tmin| < |RHmax| < |RHmed| < |Tmed| < |RHmin| < |Tmax| < |S|$, e, com os valores numéricos de $0,081 < 0,23 < -0,52 < 0,56 < 0,58 < 0,66 < 0,83$, respectivamente.

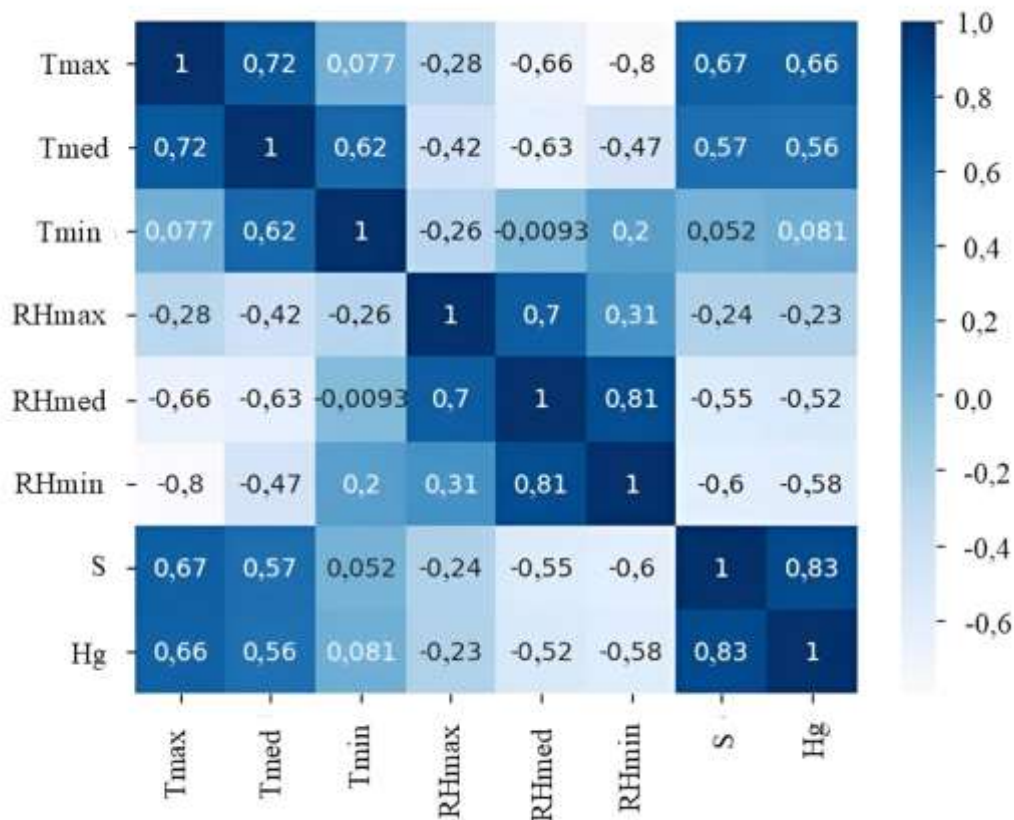


Figura 4.4. Correlação de Pearson (r) entre as variáveis meteorológicas obtidas pelas 20 EMAs e EMCs localizadas no bioma Amazônico brasileiro.

Na Tabela 4.3 é apresentado os valores médios das principais variáveis meteorológicas e empíricas a fim de caracterizar as condições ambientais onde estão inseridas as EMAs e EMCs. As condições climáticas do bioma Amazônico são complexas para a modelagem da radiação global, pois nessa região a precipitação pluviual média está entre 1.616 ± 100 e 3.205 ± 129 mm ano⁻¹, que acaba interferindo diretamente no balanço de radiação e energia, e conseqüentemente em todas as outras variáveis.

Foram avaliadas as capacidades preditivas das técnicas MLP (Figura 4.5) e SVM (Figura 4.6) para 40 diferentes combinações de variáveis de entrada, em 20 estações meteorológicas, no bioma Amazônico brasileiro. O emprego do MLP, com apenas S, So e Ho como variáveis de entrada, as médias de R², MAE, MBE, RMSE e índice “d” foram de 0,7986, 1,52 MJ m⁻² dia⁻¹, 0,013 MJ m⁻² dia⁻¹, 1,95 MJ m⁻² dia⁻¹ e de 0,9394, respectivamente.

Nas possibilidades de estimativas quando existe a disponibilidade apenas de dados diários da temperatura do ar, o emprego da combinação Tmax, Tmed, Tmin, So, Ho e propicia a melhor estimativa valores de R², MAE, MBE, RMSE e índice “d” de 0,6864, 1,92 MJ m⁻² dia⁻¹, 0,0004 MJ m⁻² dia⁻¹, 2,46 MJ m⁻² d⁻¹ e 0,8966 respectivamente. Na disponibilidade de dados diários apenas de umidade relativa do ar, a combinação RHmax, RHmed, RHmin, So, Ho resulta nas melhores estimativas.

As combinações a partir da inclusão da insolação (S) (a partir da combinação 25) melhoraram o desempenho estatístico, independentemente do número de variáveis de entrada associadas à temperatura e umidade relativa do ar. Notadamente, têm-se a percepção de que aumentar o número de variáveis de entrada pode-se melhorar o desempenho das técnicas de ML nas estimativas, o que foi observado com o MLP; sendo que entre as combinações 25 e 40, os valores médios de R², MAE, MBE, RMSE e índice “d” foram de 0,84, 1,4 MJ m⁻² dia⁻¹, 0,02 MJ m⁻² dia⁻¹, 1,70 MJ m⁻² dia⁻¹ e 0,95, respectivamente.

Tabela 4.3. Médias diárias anuais das variáveis meteorológicas e empíricas para as 20 estações meteorológicas avaliadas no bioma Amazônico.

Cidades	Hg	Ho	S	Tmax	Tmed	Tmin	RHmax	RHmed	RHmin	Prec.
1	17,17±4,80	36,23±3,40	5,58±3,11	31,29±2,84	25,60±2,04	21,68±1,96	91,75±9,71	78,42±12,52	57,89±15,46	2954±139
2	19,86±5,28	36,12±1,35	6,95±3,23	31,76±1,65	27,54±1,22	23,97±0,73	92,67±2,44	76,56±1,22	55,95±9,10	2100±145
3	17,17±5,23	35,99±1,34	4,77±3,12	32,02±2,29	26,34±1,23	22,76±1,18	96,16±2,42	83,88±6,11	58,88±10,01	2443±72
4	15,64±4,25	36,36±2,55	3,94±2,70	31,55±2,27	25,92±1,48	22,24±1,39	86,59±14,76	70,16±14,29	45,52±16,55	1952±75
5	16,12±5,09	36,05±1,98	5,78±3,33	31,52±2,24	27,24±1,44	24,01±0,98	92,72±2,69	79,57±6,57	59,88±10,56	2339±104
6	17,15±3,84	35,76±2,95	5,24±3,30	32,75±2,10	26,70±1,30	22,57±1,51	94,28±1,43	78,86±5,96	51,90±10,34	2230±103
7	16,34±5,04	35,91±2,03	5,52±3,23	32,30±2,21	27,74±1,64	24,32±1,22	91,58±6,38	75,86±9,16	54,41±11,12	2206±99
8	17,52±5,41	35,88±1,84	6,17±3,41	31,29±2,07	27,15±1,43	24,24±1,09	92,66±3,67	81,09±6,72	62,05±9,19	2343±110
9	15,22±4,76	36,17±1,30	4,73±2,81	31,30±2,23	26,41±1,45	23,14±1,19	93,13±5,30	81,46±7,99	59,18±10,41	2867±46
10	19,13±4,19	35,95±3,96	6,03±3,04	32,35±2,81	25,41±1,63	20,16±2,11	91,69±8,05	72,04±15,78	44,38±16,82	1952±132
11	15,09±3,59	36,04±1,55	6,48±2,75	32,67±1,35	27,27±1,09	23,56±0,65	93,22±2,37	78,49±5,75	54,95±7,22	3205±129
12	20,16±3,78	35,91±1,79	7,57±2,59	32,47±1,21	27,75±1,13	24,23±1,02	88,92±4,04	74,36±6,15	53,30±6,81	2230±137
13	18,64±4,46	35,79±3,26	6,96±3,26	33,54±2,75	26,83±1,69	21,60±2,12	90,66±6,24	70,50±12,26	43,56±15,08	1686±104
14	18,75±4,71	36,03±2,25	6,24±3,18	32,67±2,17	27,58±1,46	23,85±0,96	86,22±10,94	74,87±7,16	60,38±12,95	2069±95
15	18,25±3,87	35,82±2,57	6,36±3,10	32,26±1,95	26,59±1,14	22,40±1,37	93,31±2,82	76,53±7,75	50,78±11,48	1885±123
16	20,61±4,19	36,13±1,71	7,53±2,79	31,66±1,69	27,54±1,29	23,97±1,05	87,92±5,38	75,30±6,98	55,21±8,88	1661±104
17	16,64±4,52	36,21±2,31	6,70±3,21	33,08±2,45	26,84±1,46	22,74±0,78	92,77±3,72	78,22±8,71	52,84±11,51	2572±107
18	19,82±4,30	35,96±1,38	6,89±3,55	30,94±0,95	27,71±1,04	25,34±1,51	86,30±6,78	76,98±6,03	64,05±5,21	2093±74
19	16,95±3,48	36,06±1,99	6,22±2,81	31,43±1,68	26,73±1,15	23,36±0,94	94,25±4,29	78,42±7,70	56,01±9,19	2400±157
20	19,35±4,35	35,99±1,77	6,49±2,87	33,51±2,22	27,83±1,56	23,70±1,07	86,69±7,71	68,54±10,17	45,03±10,41	1616±100

Radiação global (Hg, MJ m⁻² d⁻¹), radiação extraterrestre (Ho, MJ m⁻² d⁻¹), coeficiente de transmissividade (Hg/Ho), insolação (S, horas), fotoperíodo (So, horas), razão de insolação (S/So), temperatura máxima (Tmax, °C), temperatura média (Tmed °C), temperatura mínima (Tmin, °C), umidade relativa máxima (RHmax, %), umidade relativa média (RHmed, %), umidade relativa mínima (RHmin) e precipitação pluvial (Prec., mm ano⁻¹).

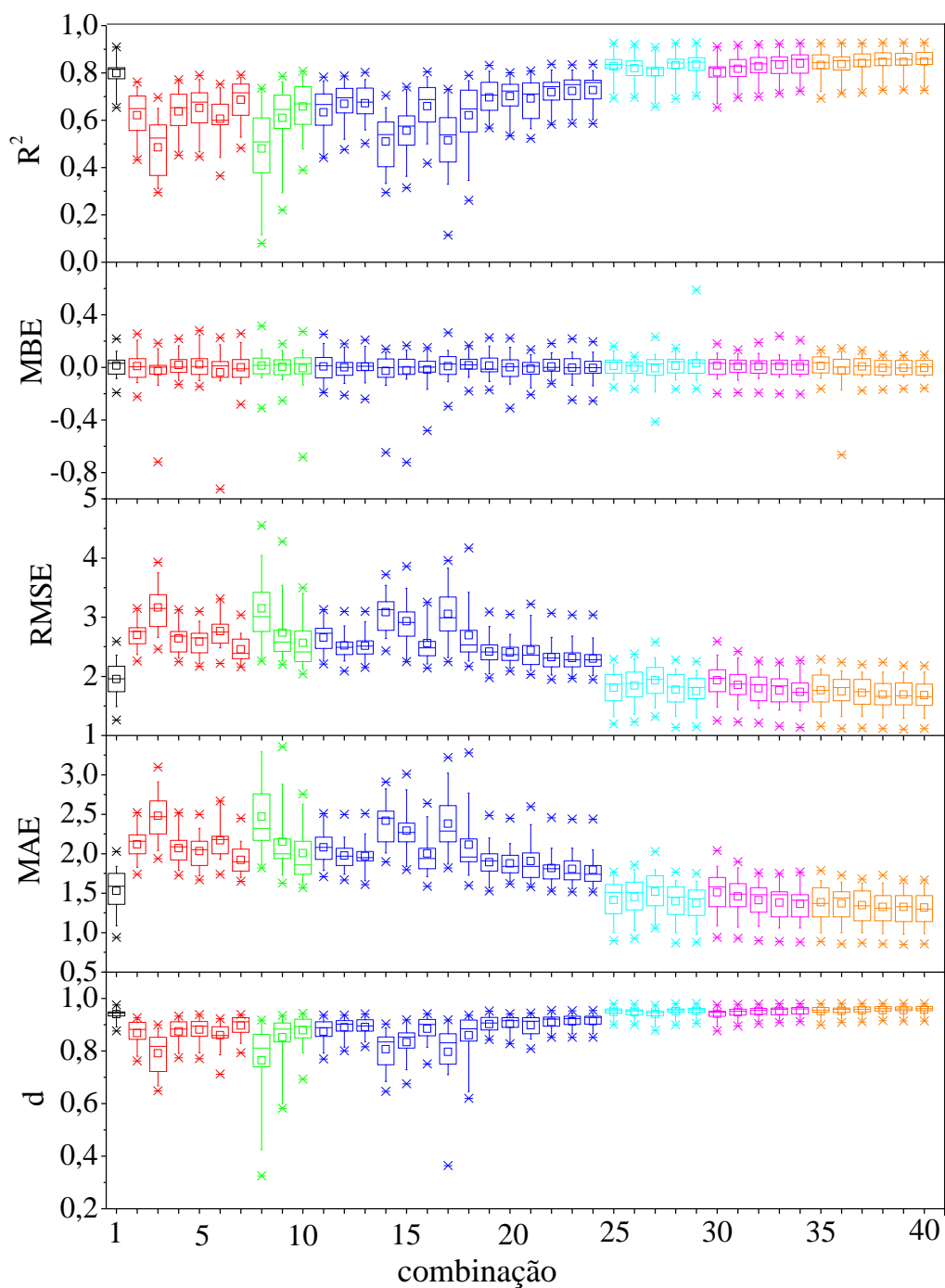


Figura 4.5. Boxplot dos coeficientes de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE), erro absoluto médio (MAE) e coeficiente de Willmott (d), para o MLP em 40 combinações de variáveis de entrada, em 20 estações meteorológicas do bioma Amazônico brasileiro. (As diferentes cores representam os agrupamentos das combinações).

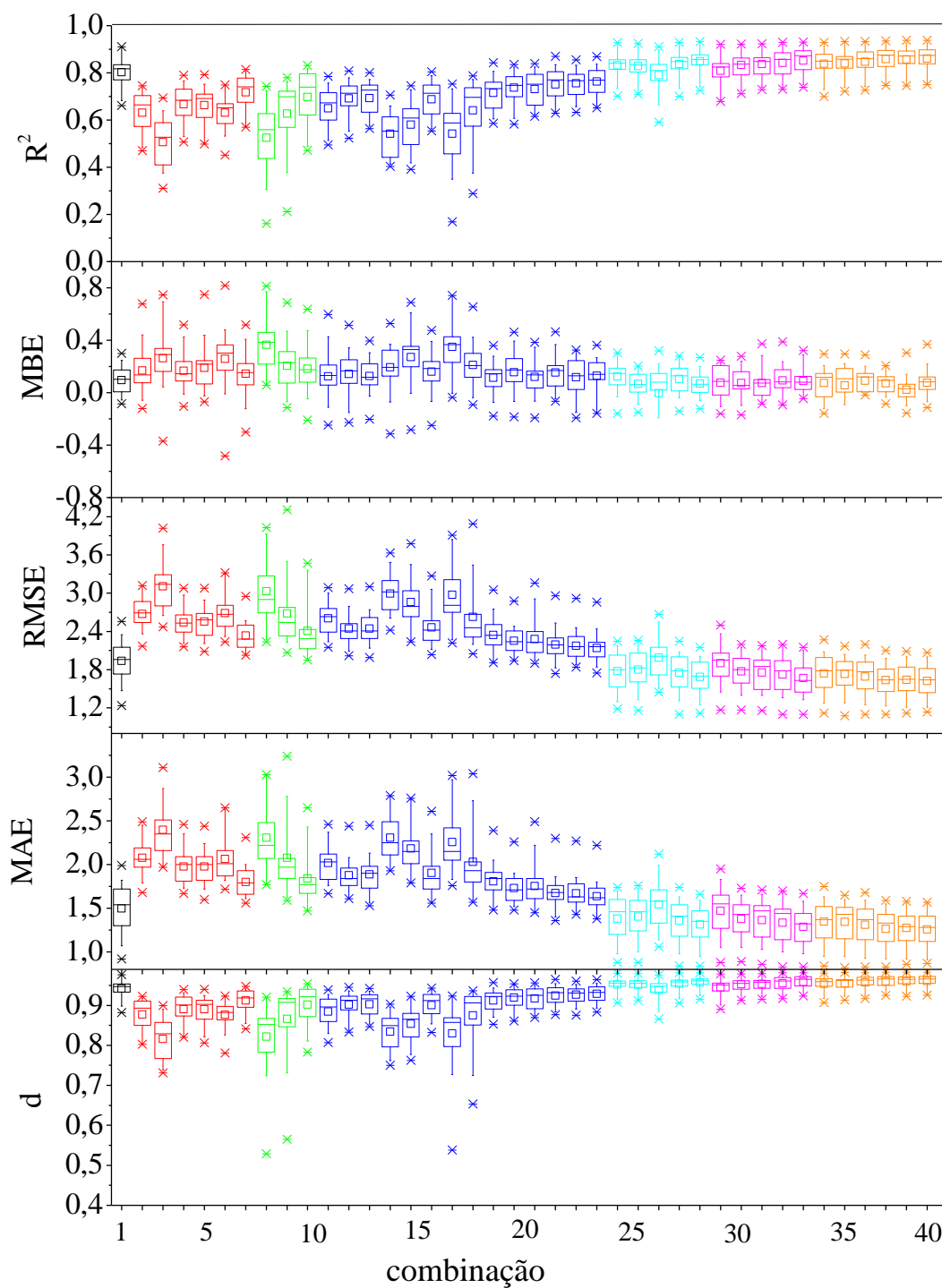


Figura 4.6. Boxplot dos coeficientes de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro relativo médio (RMSE), erro absoluto médio (MAE) e coeficiente de Willmott (d), para o SVM em 40 combinações de variáveis de entrada, em 20 estações meteorológicas do bioma Amazônico brasileiro. (As diferentes cores representam os agrupamentos das combinações).

O SVM apresentou capacidade preditiva semelhante ao MLP para todas as combinações e agrupamentos, com melhores valores dos indicativos R^2 , MAE, RMSE e índice “d” de Willmott. Por exemplo, para a combinação com S, So e Ho como variáveis de entrada, os valores de R^2 , MAE, MBE, RMSE e “d” foram de 0,8024, 1,49, 0,096 MJ m⁻² dia⁻¹, 1,93 MJ m⁻² dia⁻¹ e de 0,9421, respectivamente. Em geral, apenas nos desvios relativos (MBE) que ocorrem a presença de super ou subestimativas, o MLP propiciou menores valores, quando comparado ao SVM.

Apenas as sete combinações mais representativas (1, 7 e 40) foram selecionadas para avaliação da dispersão (Figuras 4.7 e 4.8) entre os valores estimados e medidos. Essa comparação é apresentada para estações meteorológicas de Boa Vista (latitude 2,85° - localizada no extremo norte), Manaus (latitude -3,81° - região central) e Sinop (latitude -11,98° - extremo sul), propiciando assim, uma análise espacial abrangente das características geográficas e meteorológicas inseridas no bioma Amazônico e da interferência dessas condições regionais na análise da modelagem por ML. Nesse caso, a radiação global (Hg) foi dividida em função do coeficiente de transmissividade atmosférica (Kt), em quatro intervalos (destacados em diferentes cores): $0 \leq Kt < 0,35$ (preto), $0,35 \leq Kt < 0,55$ (vermelho), $0,55 \leq Kt < 0,65$ (verde) e $Kt \geq 0,65$ (azul), que correspondem as condições de céu nublado, parcialmente nublado com predominância de radiação difusa, parcialmente aberto com predominância de radiação direta e céu aberto, respectivamente, conforme Escobedo et al. (2009).

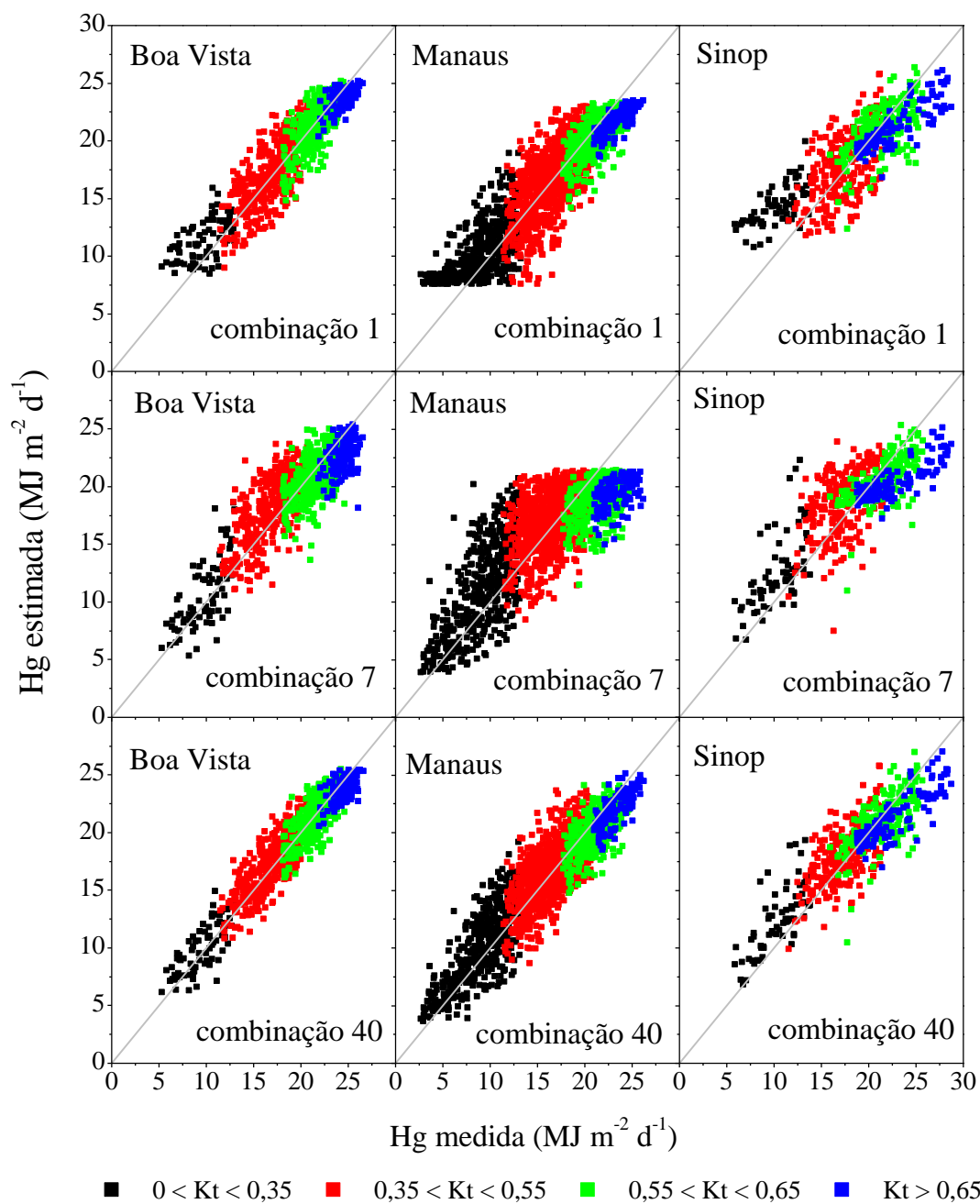


Figura 4.7. Dispersão entre valores medidos e estimados de Hg pelo MLP em diferentes condições de cobertura de céu, nas estações meteorológicas de Boa Vista, Manaus e Sinop, na Amazônia brasileira.

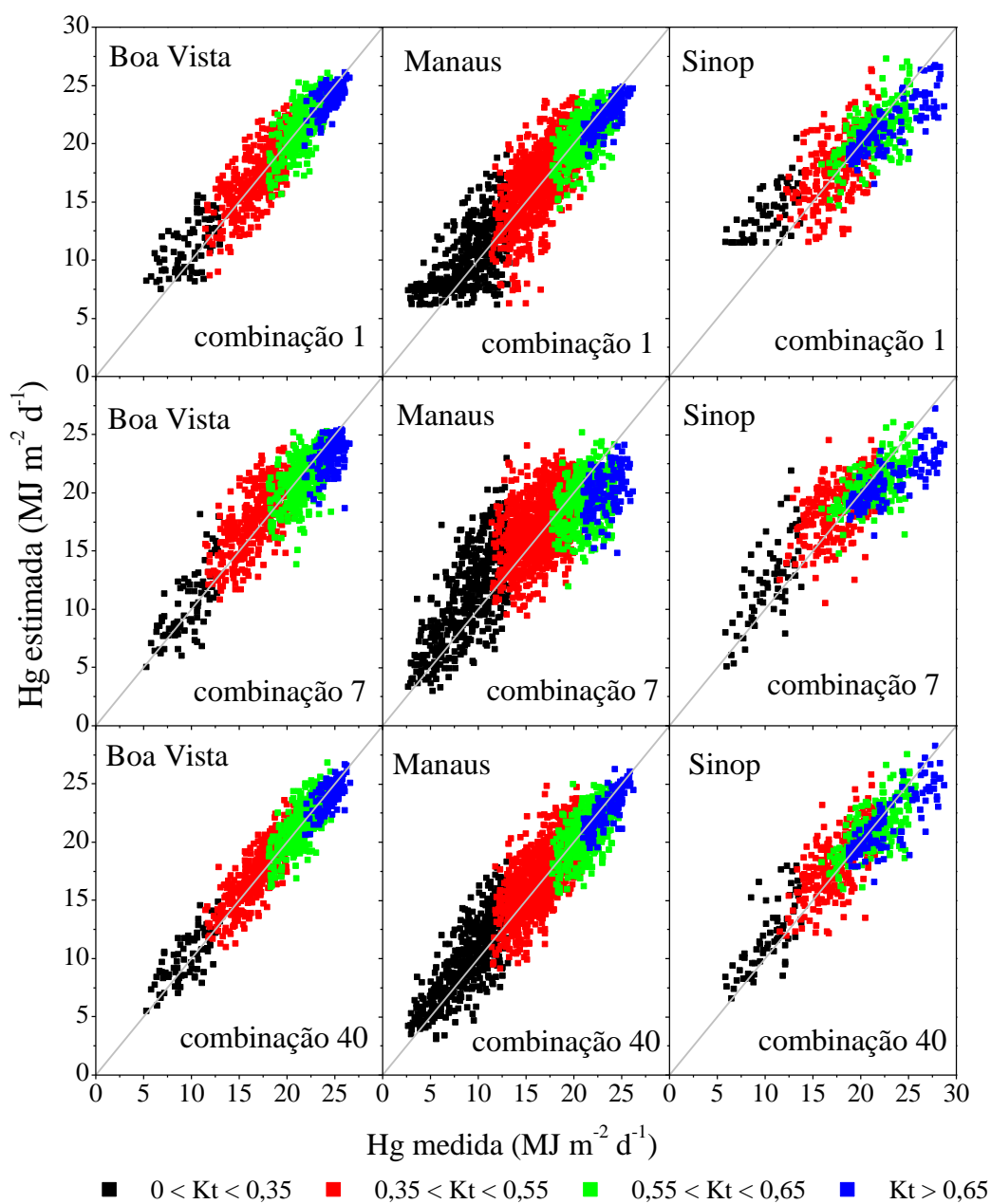


Figura 4.8. Dispersão entre valores medidos e estimados de Hg pelo SVM em diferentes condições de cobertura de céu, nas estações meteorológicas de Boa Vista, Manaus e Sinop, na Amazônia brasileira.

A capacidade preditiva do MLP e SVM (Figura 4.7 e 4.8) mostrou-se variável entre as três estações meteorológicas avaliadas, com maiores dispersões para a estação de Manaus. Outro ponto importante, é que as estimativas de Hg foram mais próximas das medidas para condições de céu nublado ou aberto. As combinações híbridas, que consideram as variáveis de entrada a insolação, temperatura e umidade relativa do ar

apresentam melhores estimativas de Hg para céu parcialmente nublado com predominância de radiação difusa, em dias com alta transmissividade atmosférica houve redução no espalhamento dos valores da radiação estimada, tanto para o MLP quanto SVM.

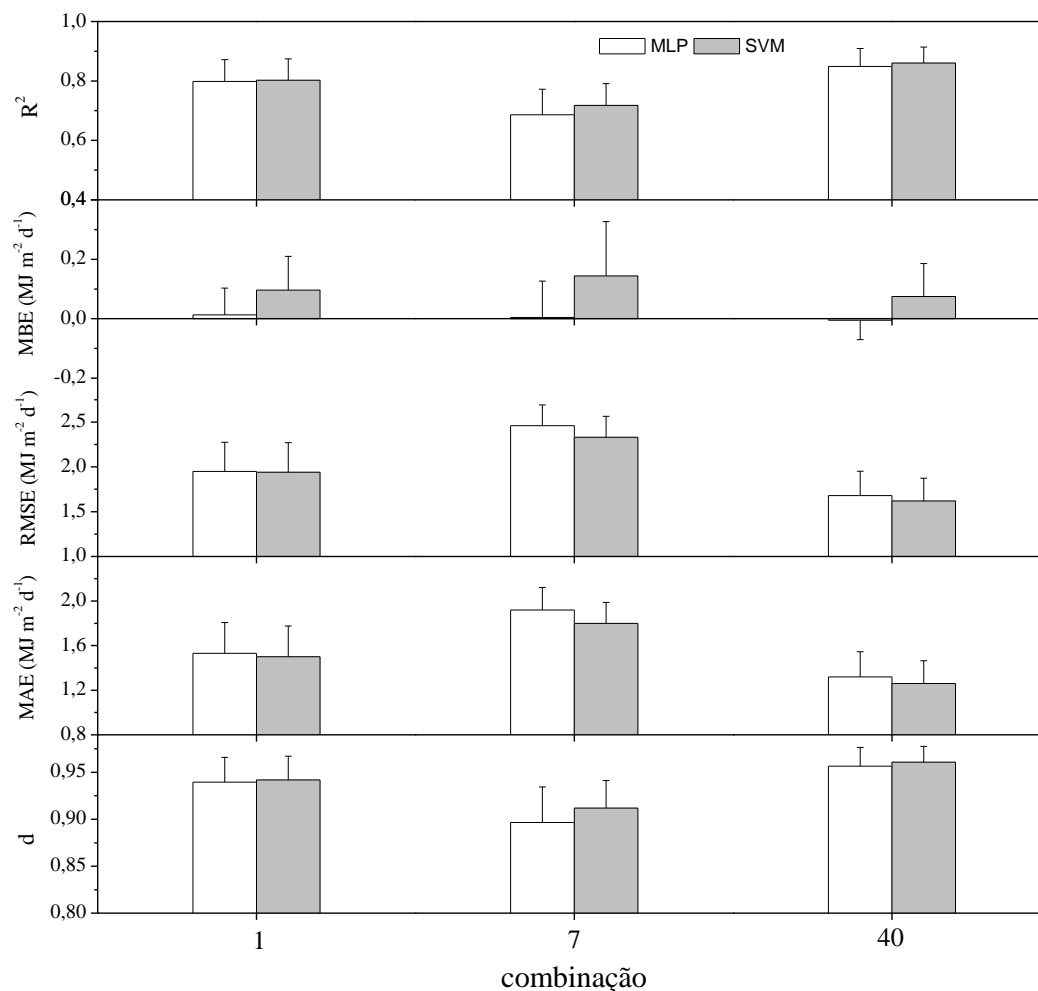


Figura 4.9. Valores médios do coeficiente de determinação (R^2), erro relativo médio (MBE), raiz quadrada do erro médio (RMSE), erro médio absoluto (MAE) e índice de concordância de Willmott (d), para o MLP (branco) e SVM (cinza), em diferentes combinações de variáveis de entrada (1, 7 e 40).

Para as três melhores combinações (1, 7 e 40), foi comparado o desempenho entre o MLP e SVM (Figura 4.9). O desempenho do SVM foi superior em todas as combinações, com R^2 variando de 0,70 a 0,86, MAE de 1,26 a 1,90 $\text{MJ m}^{-2} \text{dia}^{-1}$, RMSE de 1,62 a 2,40 $\text{MJ m}^{-2} \text{dia}^{-1}$ e índice “d” variando de 0,91 a 0,96. Dentre os

índices estatísticos avaliados de qualidade dos ajustes, apenas o MBE para o MLP apresentou valores médios melhores, com variação de $-0,008$ a $0,029 \text{ MJ m}^{-2} \text{ dia}^{-1}$.

Com o intuito de entender os avanços que as técnicas MLP e SVM podem gerar nas estimativas de Hg, foram realizadas comparações com modelos empíricos que adotam as mesmas variáveis de entrada, para as estações meteorológicas supracitadas de Boa Vista-RR, Manaus-AM e Sinop-MT. Dessa forma, considerando os modelos avaliados no capítulo 1 desta tese, têm-se as seguintes comparações: i) para insolação (S) - MLP1, SVM1 e modelo 10 (S, So e Ho); ii) para temperatura do ar – MLP7, SVM7 e modelo 45 (Tmax, Tmed, Tmin, So e Ho); iii) combinações híbridas – MLP28, SVM28 e modelo 62 (Tmax, Tmin, S, So e Ho). Não foram observadas diferenças significativas no desempenho entre as estimativas dos modelos simplificados 10, 45 e 62 quando comparados com MLP e SVM (Figura 4.10).

Não houve diferença, na comparação entre os modelos empíricos e as técnicas de ML, com a frequência do erro relativo (Figura 4.11) acumulado até o valor de $2,0 \text{ MJ m}^{-2} \text{ d}^{-1}$ para o modelo 10, MLP1 e SVM1 que foi de 76, 76 e 76%, modelo 45, MLP7 e SVM7 com 71, 71 e 71% e o modelo 62, MLP28 e SVM28 com 80, 82 e 83%.

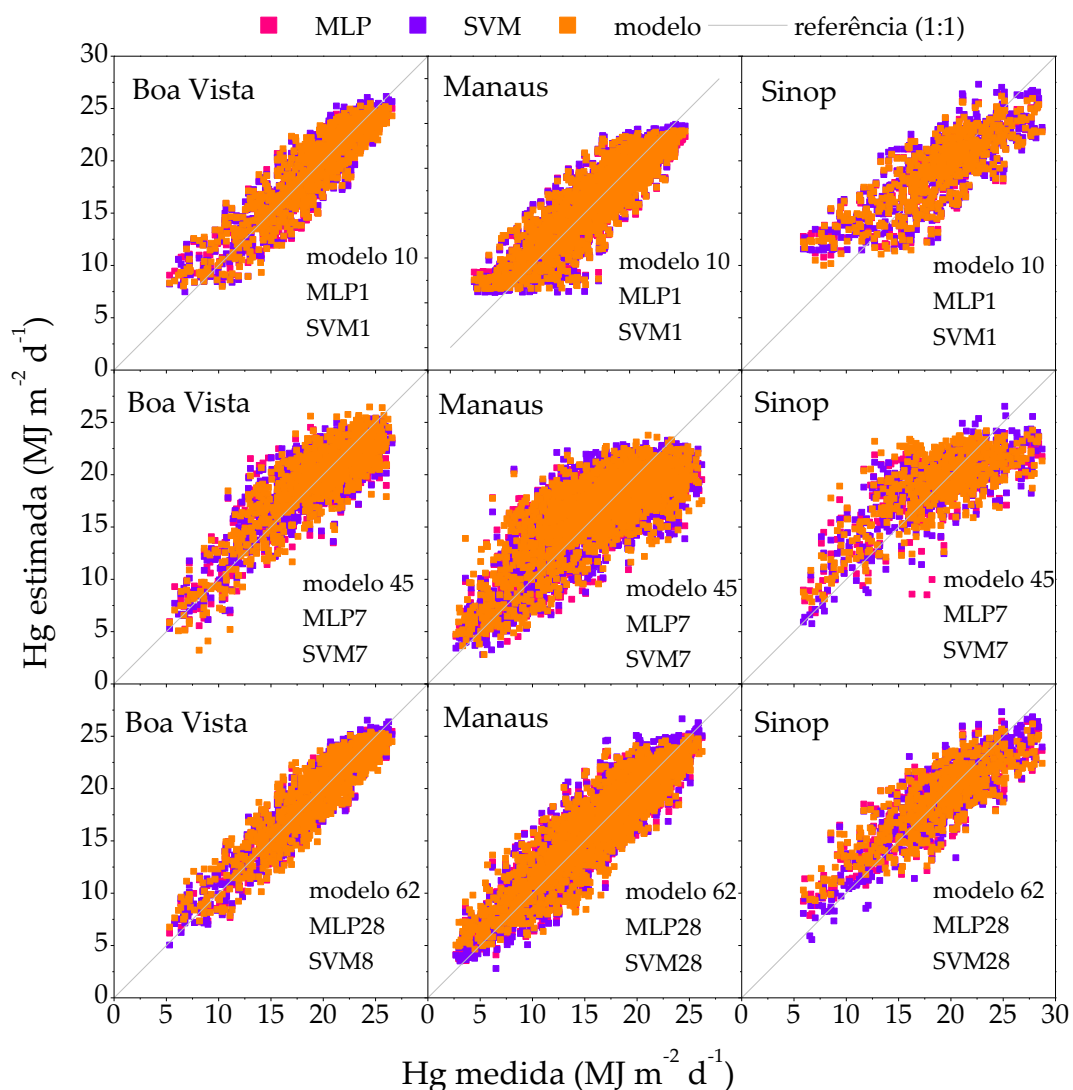


Figura 4.10. Dispersão entre radiação global medida e radiação global estimada por MLP, SVM e modelos empíricos simplificados, para as estações meteorológicas de Boa Vista-RR, Manaus-AM e Sinop-MT, considerando diferentes variáveis de entrada dos modelos.

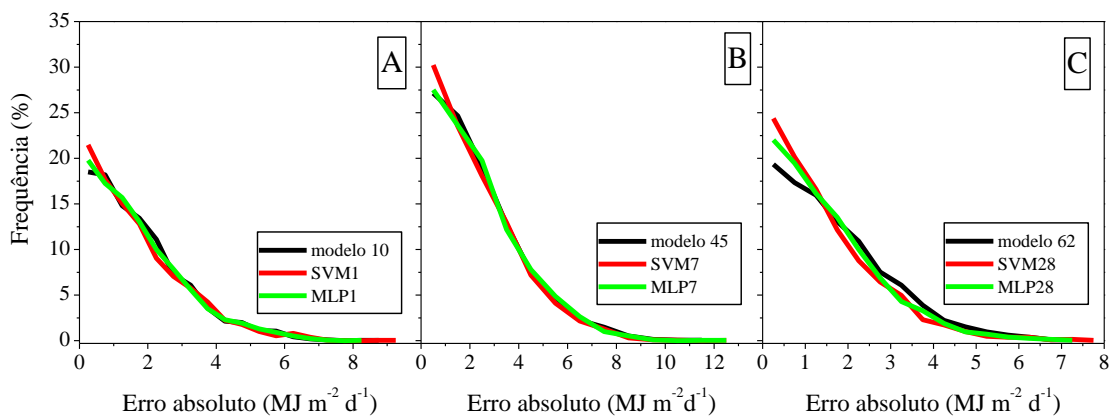


Figura 4.11. Frequência do erro absoluto na estimativa da radiação global empregando modelo empírico e técnicas de ML com a insolação como variável de entrada (A), temperatura do ar (B) e combinação híbrida (C) como variável entrada para as estações meteorológicas de Boa Vista, Manaus e Sinop.

4.4 DISCUSSÃO

A modelagem da Hg incidente na superfície terrestre é complexa, pois esse elemento e fator meteorológico é influenciado pela atmosfera, que é dinâmica e composta por diversos elementos, como gases, poeiras, vapor d'água e nuvens (NAWAB et al., 2023). Esses componentes atmosféricos interagem com diferentes comprimentos de onda da radiação e geram processos como espalhamentos, reflexão e absorção. Segundo Li et al. (2023), a previsão da Hg se torna mais difícil à medida que a transmissividade atmosférica (K_t) diminui, ou seja, condição de dias nublados/chuvosos ou sob condições de aumento das concentrações de materiais particulados em suspensão.

Sabendo que as condições geográficas locais ou regionais influenciam diretamente na sazonalidade e na distribuição espacial das variáveis meteorológicas, a seleção de diferentes variáveis de entrada aplicadas na estimativa da Hg deve ser criteriosa e avaliada/calibrada para diferentes condições locais (HE et al. 2020). As pesquisas envolvendo modelagem micrometeorológicas, tanto com modelos empíricos como com técnicas de ML, devem ser sustentadas em variáveis com medição amplamente difundida, de baixo custo e com sensores de fácil implementação (JIA et al. 2022). Nesse caso, quanto maior o número de pesquisas desenvolvidas com essa temática e abordagem, melhor será a capacidade preditiva das técnicas de ML (ZHOU et al. 2021; WOLDEGIYORGIS et al. 2023).

Quanto maior o número de variáveis de entrada nas técnicas de ML, considerando as mesmas unidades de medidas e partições temporais (instantâneas, horárias ou diárias) [como por exemplo - variações da temperatura do ar - T_{max} , T_{med} , T_{min} , $T_{max} - T_{min}$ e T_{max}/T_{min}], espera-se que ocorra melhorias na capacidade preditiva e nos indicadores estatísticos de desempenho (BELLIDO-JIMÉNEZ et al. (2021), sendo que as técnicas de ML são metodologias inicialmente desenvolvidas para resolver problemas complexos, não-lineares e com grande número

de variáveis (BOUNOUA et al. 2021), pois essa condição é bem recorrente em regiões de clima tropical como o bioma Amazônico, com alta precipitação pluvial anual (ALVARES et al. 2013), que geram dinâmicas atmosféricas distintas e podem inferir ruídos e interferir na capacidade preditiva da Hg. A vantagem de avaliar diferentes combinações de variáveis de entrada é que quando não há disponibilidade de uma determinada variável, pode ser escolhida outra combinação que contemple as variáveis disponíveis e que apresentem redução de erros de estimativas (HE et al., 2020).

Os resultados mostraram que dependendo da combinação utilizada (Figuras 4.5 e 4.6), a capacidade preditiva tanto do MLP como SVM fica seriamente comprometida, com a piora nos desempenhos estatísticos. A utilização de RHmax e Tmin (juntas ou separadamente) deve ser evitada, visto que no ciclo diurno/noturno, nos instantes de ocorrência de RHmax e Tmin não ocorre incidência de radiação global, e, por isso, essas variáveis apresentam baixa correlação significativa com Hg. Esses resultados corroboram com He et al. (2020), que avaliando o SVM na estimativa da Hg em 80 cidades localizadas na China, com diferentes combinações de entrada, verificaram que, no geral, o aumento do número de variáveis também melhorou o desempenho das estimativas, porém algumas variáveis quando adicionadas não geraram melhores desempenhos estatísticos. Segundo esses autores, Tmin é a variável com menor correlação com a Hg quando comparadas com as variações da temperatura do ar (amplitude térmica, Tmax, Tmed e Tmin). Para Kaba et al. (2018), as técnicas de ML, com o aumento do número de variáveis de entrada também geram melhorias nas estimativas de Hg, porém quando se emprega Tmin em diferentes combinações, não são observadas melhorias significativas.

Numericamente, os melhores desempenhos estatísticos foram obtidos quando se utilizou a combinação híbrida com todas as variáveis de entrada (RHmax, RHmed, RHmin, Tmax, Tmed, Tmin, S, So, Ho); todavia, são valores semelhantes aos observados para as combinações vinculadas apenas à insolação e a temperatura do ar (Tmax e Tmed). Huang et al. (2022), em três diferentes condições climáticas, com apenas uma variável meteorológica, obtiveram melhores estimativas com Tmed; já com os modelos híbridos, esses mesmos autores observaram que a combinação entre Tmed, velocidade do vento, umidade relativa e precipitação pluvial geram boas estimativas de Hg.

A amplitude dos valores dos indicadores estatísticos quando se analisa as diferentes estações meteorológicas, está relacionada com a cobertura de nuvens e sazonalidade das precipitações, vegetação e proximidade à grandes superfícies de águas livres, que podem aumentar a umidade relativa do ar. Além disso, as mudanças ambientais propiciadas por ações antrópicas próximas aos pontos de medição (estações) como atividade industrial e queimadas que emitem material particulado (MARQUES et. al. 2024), podem interferir no balanço de radiação e energia. Quando se compara as três estações em diferentes latitudes da Amazônia brasileira (Figura 4.7 e 4.8), observa-se que o desempenho das técnicas de ML é dependente das condições climáticas locais.

As técnicas de ML podem estimar a Hg com boa precisão em uma determinada região, porém essa mesma técnica, quando utilizada em outras regiões pode apresentar piores estimativas quando comparado com outros modelos. Isso acontece também quanto às variáveis de entrada, visto que para uma determinada região, por exemplo, a insolação gera melhores estimativas quando comparada com a temperatura do ar, e, em outras regiões pode ocorrer o inverso (GÜREL et al. 2023). Bounoua et al. (2021) mostraram que o desempenho estatístico do MLP foi diferente em cinco cidades avaliadas e relacionaram esse comportamento com a variabilidade nas condições climáticas e qualidade das medições. Segundo Agbulut et al. (2021) não existe um técnica de ML que consiga um bom desempenho em todas as regiões geográficas e climáticas do mundo, pois são diretamente dependentes das condições locais, tamanho do conjunto de dados, características geográficas e principalmente dos hiperparâmetros, que devem ser fornecidos aos modelos como número de neurônios, camada oculta, normalização dos valores de entrada, e, que muitas vezes são subjetivos e demandam testes e avaliações de indicadores estatísticos. Para Gürel et al. (2023), analisar os parâmetros dos modelos, técnicas de ML e conjunto de dados para uma região de interesse é fundamental para se ter boas estimativas de qualquer variável ambiental.

Observa-se que as combinações acima de duas variáveis de entrada propiciam melhores estimativas de Hg, sendo que o MLP e SVM se tornam limitados quando avaliados com poucas variáveis de entrada, melhorando quando empregado a combinação envolvendo insolação em conjunto com temperatura do ar ou umidade

relativa do ar. Husain & Khan (2021) avaliaram 12 modelos de ML com diferentes combinações de entrada, em clima subtropical úmido na Índia, e as combinações acima de duas variáveis como temperatura do ar, umidade relativa do ar e insolação melhoraram os desempenhos de MLP e SVM, porém, segundo Nawab et al. (2023), as variáveis que mais afetam na melhora do desempenho das técnicas de ML são Tmax, Tmin, ΔT , RH, Kt e precipitação pluvial.

A capacidade preditiva da Hg empregando o SVM em clima tropical foi superior quando comparado ao MLP (Figura 4.9). Na literatura, observa-se que esse comportamento é dependente das condições locais (QUEJ et al. 2017; FAN et al. 2019; MARZOUQ et al. 2019; AGBULUT et al. 2021; BOUNOUA et al. 2021; HUSAIN & KHAN, 2021; JIA et al. 2022; LIU et al. 2022; NEMATCHOUA et al. 2022; KÜÇÜKTOPÇU et al. 2024). Para Bellido-Jiménez et al. (2021), os modelos MLP são melhores em climas áridos e semiáridos, enquanto o SVM é melhor em climas úmidos. Por conseguinte, He et al. (2020) destacam que o SVM é o método mais empregado para resolver problemas com dados de alta dimensionalidade e não linear, pois consegue contornar mais facilmente dados com algum tipo de ruído.

Nas condições brasileiras, em Botucatu-SP, Silva et al. (2017) analisando o coeficiente de correlação (r) e índice de Willmott (d) na validação do SVM e MLP encontraram para anos típicos com razão de insolação (S/S_o) e H_o como variáveis de entrada, que o SVM apresentou R^2 de 0,96 e 0,98 e foi melhor do que MLP (R^2 de 0,924 e 0,910); porém, com a inclusão de S/S_o , H_o , Tmax e Tmin como variáveis de entrada, os indicadores estatísticos melhoraram significativamente para ambas as técnicas de ML.

Os estudos relacionados às estimativas de Hg mostram que a maioria das técnicas de ML são mais precisas quando comparados aos modelos empíricos (FENG et al. 2019; NAWAB et al. 2023). Todavia, essa condição depende da variável de entrada (Figura 4.10), associado com a variação sazonal das precipitações e da transmissividade atmosférica (JIA et al. 2022; LI et al. 2023). Antonopoulos et al. (2019) comparando diferentes metodologias de estimativa de Hg e diferentes combinações de entrada das técnicas de ML para a Grécia, observaram que o método de regressão linear múltipla (MLR) apresentou o melhor desempenho com a combinação H_o , ΔT , $\Delta T^{0,5}$ e RHmed, seguidos do modelo empírico de Hargreaves &

Samani; nesse caso, ambos os métodos apresentaram melhores desempenhos estatísticos quando comparados a rede neural artificial (ANN) com as mesmas variáveis de entrada.

Existem vantagens e desvantagens quando avaliadas às diferentes metodologias, sendo a ampla utilização dos modelos empíricos potencializadas pela sua simplicidade e precisão, porém, são viáveis apenas em regiões com condições climáticas específicas, pois os parâmetros dos modelos são fixos (coeficientes) ou calibrados localmente. Por sua vez, os modelos de ML, são mais precisos quando envolvem problemas não lineares, séries temporais extensas e estes suportam dezenas, centenas ou milhares de variáveis de entrada; porém, a otimização dos hiperparâmetros quando não levados em consideração podem ser a limitação mais significativa nestas metodologias. As condições climáticas são de natureza não linear (WOLDEGIYORGIS et al. 2023), aliado ao fato de que as mudanças climáticas estão se intensificando em diversas regiões brasileiras e outros países, e geram mudanças nos balanços de radiação e energia a nível local e regional, demandando assim, a necessidade de avaliações periódicas das técnicas de ML e recalibração de modelos empíricos.

Os resultados encontrados em uma determinada região, podem ser ampliados para condições climáticas semelhantes (BELLIDO-JIMÉNEZ et al. 2021; BOUNOUA et al. 2021). Nesse caso, para regiões de clima tropical, se atentando principalmente as condições de Kt locais. Recomenda-se o emprego das variáveis insolação (quando disponível), temperatura do ar e combinações híbridas.

4.5 CONCLUSÃO

O aumento do número de variáveis de entrada melhorou significativamente o desempenho das técnicas de Machine Learning, sendo que a melhor combinação envolve as variáveis meteorológicas com a insolação que reduz o espalhamento com condições de alta transmissividade atmosférica e a temperatura do ar que reduz a dispersão dos valores estimados em condições de baixa transmissividade atmosférica.

Máquina de Vetor e Suporte (SVM) apresentou desempenho superior na estimativa da radiação global, quando comparado ao Multi Layer Perceptron (MLP) em todas as estações meteorológicas avaliadas. Comparando as diferentes metodologias, foi obtido o desempenho estatístico próximo da estimativa da radiação global, quando se compara as técnicas de Machine Learning em relação aos modelos empíricos com emprego da insolação e temperatura do ar no bioma Amazônico brasileiro.

4.6 REFERÊNCIAS BIBLIOGRÁFICAS

ABSOLAR. Panorama da solar fotovoltaica no Brasil e no mundo. Disponível em: <<https://www.absolar.org.br/mercado/infografico/>>. Acesso em 12 de maio de 2024.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013. DOI: 10.1127/0941-2948/2013/0507.

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. **Crop evapotranspiration guidelines for computing crop water requirements e FAO irrigation and drainage paper 56**. Rome: Food and Agriculture Organization of the United Nations; 1998.

ANTONOPOULOS, V.; PAPAMICHAIL, D. M.; ASCHONITIS, V. G.; ANTONOPOULOS, A. V. Solar radiation estimation methods using ANN and empirical models. **Computers and Electronics in Agriculture**, v. 160, p. 160-167, 2019. <https://doi.org/10.1016/j.compag.2019.03.022>.

AGBULUT, Ü.; GÜREL, A. E.; BIÇEN, Y. Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. **Renewable and Sustainable Energy Reviews**, v. 135, p. 110114, 2021. <https://doi.org/10.1016/j.rser.2020.110114>.

BELLIDO-JIMÉNEZ, J.; GUALDA, J. E.; GARCÍA-MARÍN, A. P. Assessing new intra-daily temperature-based machine learning models to outperform solar radiation predictions in different conditions. **Applied Energy**, v. 298, p. 117211, 2021. <https://doi.org/10.1016/j.apenergy.2021.117211>.

BOUNOUA, Z.; CHAHIDI, L. O.; MECHAQRANE, A. Estimation of daily global solar radiation using empirical and machine-learning methods: A case study of five Moroccan locations. **Sustainable Materials and Technologies**, v. 28, p. 261, 2021. <https://doi.org/10.1016/j.susmat.2021.e00261>.

ESCOBEDO, J. F.; GOMES, E. N.; OLIVEIRA, A. P.; SOARES, J. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. **Applied Energy**, v. 86, n. 3, p. 299-309, 2009. <https://doi.org/10.1016/j.apenergy.2008.04.013>.

FAN, J.; WU, L.; ZHANG, F.; CAI, H.; ZENG, W.; WANG, X.; ZOU, H. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. **Renewable and Sustainable Energy Reviews**, v. 100, p. 186-212, 2019. <https://doi.org/10.1016/j.rser.2018.10.018>.

FENG, Y.; GONG, D.; ZHANG, Q.; JIANG, S.; ZHAO, L.; CUI, N. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. **Energy Conversion and Management**, v. 198, n. 15, 2019. <https://doi.org/10.1016/j.enconman.2019.111780>.

GÜREL, A. E.; AGBULUT, Ü.; BAKIR, H.; ERGÜN, A.; YILDIZ, G. A state of art review on estimation of solar radiation with various models. **Heliyon**, v. 9, n. 2, 2023. <https://doi.org/10.1016/j.heliyon.2023.e13167>.

HE, C.; LIU, J.; XU, F.; ZHANG, T.; CHEN, S.; SUN, Z.; ZHENG, W.; WANG, R.; HE, L.; FENG, H.; YU, Q.; HE, J. Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods. **Energy Conversion and Management**, v. 220, n. 15, 2020. <https://doi.org/10.1016/j.enconman.2020.113111>.

HUANG, H.; BAND, S.; KARAMI, H.; EHTERAM, M.; CHAU, K-W.; ZHANG, Q. Solar radiation prediction using improved soft computing models for semi-arid, Slightly-arid e humid climates. **Alexandria Engineering Journal**, v. 61, p. 10631-10657, 2022. <https://doi.org/10.1016/j.aej.2022.03.078>.

HUSAIN, S.; KHAN, U. A. Machine Learning models to predict diffuse solar based on diffuse fraction and diffusion coefficient models for humid-subtropical climatic zone of India. **Cleaner Engineering and Technology**, v. 5, 2021. <https://doi.org/10.1016/j.clet.2021.100262>.

JIA, D.; YANG, L.; LV, T.; LIU, W.; GAO, X.; ZHOU, J. Evaluation of machine learning models for predicting daily global and diffuse solar radiation under different weather/pollution conditions. **Renewable Energy**, v. 187, p. 896-906, 2022. <https://doi.org/10.1016/j.renene.2022.02.002>.

KABA, K.; SARIGÜL, M.; AVCI, M.; KANDIRMAZ, H. M. Estimation of daily global solar radiation using deep learning model. **Energy**, v. 162, p. 126-135, 2018. <https://doi.org/10.1016/j.energy.2018.07.202>.

KÜÇÜKTOPÇU, E.; GEMEK, B.; SIMSEK, H. Comparative analysis of single and hybrid machine learning models for daily solar radiation. **Energy Reports**, v. 11, p. 3256-3266, 2024. <https://doi.org/10.1016/j.egyr.2024.03.012>.

LI, Y.; WANG, Y.; QIAN, H.; GOA, W.; FUKUDA, H.; ZHOU, W. Hourly global solar radiation prediction based on seasonal and stochastic feature. **Heliyon**, v. 9, n. 9, 2023. <https://doi.org/10.1016/j.heliyon.2023.e19823>.

LIU, F.; WANG, X.; SUN, F.; WANG, H. Correct and remap solar radiation and photovoltaic power in China based on machine learning models. **Applied Energy**, v. 312, n. 15, 2022. <https://doi.org/10.1016/j.apenergy.2022.118775>.

MARZOUQ, M.; BOUNOUA, Z.; FADILI, H. E.; MECHAQRANE, A.; ZENKOUAR, K. New daily global irradiation estimation model based on automatic selection of input parameters using evolutionary neural networks. **Journal of Cleaner Production**, v. 209, n. 1, p. 1105-1118, 2019. <https://doi.org/10.1016/j.jclepro.2018.10.254>.

MARQUES, A. L. F.; TEIXEIRA, M. J.; ALMEIDA, F. V.; CORRÊA, P. L. P. Neural Networks Forecast Models Comparison for the Solar Energy Generation in Amazon Basin. **IEEE Access**, v. 12, 2024. DOI: 10.1109/ACCESS.2024.3358339.

MARTIM, C. C.; PAULISTA, R. S.; CASTAGNA, D. BORELLA, D. R.; ALMEIDA, F. T.; DAMIAN, J. G. R.; SOUZA, A. P. Daily Estimates of Global Radiation in the Brazilian Amazon from Simplified Models. **Atmosphere**, v. 15, n. 11, 2024. DOI: <https://doi.org/10.3390/atmos15111397>.

NAWAB, F.; HAMID, A. S. A.; IBRAHIM, A.; SOPIAN, K.; FAZLIZAN, A.; FAUZAN, M. F. Solar irradiation prediction using empirical and artificial intelligence methods: A comparative review. **Heliyon**, v. 9, n. 6, 2023. <https://doi.org/10.1016/j.heliyon.2023.e17038>.

NEMATCHOUA, M. K.; OROSA, J. A.; AFAIFIA, M. Prediction of daily global solar radiation and air temperature using six machine learning algorithms; a case of 27 European countries. **Ecological Informatics**, v. 69, 2022. <https://doi.org/10.1016/j.ecoinf.2022.101643>.

QUEJ, V. H.; ALMOROX, J.; ARNALDO, J. A.; SAITO, L. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 155, p. 62-70, 2017. <https://doi.org/10.1016/j.jastp.2017.02.002>.

SANTOS, C. M.; TERAMOTO, É. T.; SOUZA, A.; ARISTONE, F.; IHADDADENE, R. Several models to estimate daily global solar irradiation adjustment and evaluation. **Arabian Journal of Geosciences**, v. 14, n. 4, 2021. <https://doi.org/10.1007/s12517-021-06603-8>.

SILVA, M. B. P.; ESCOBEDO, J. F.; ROSSI, T. J.; SANTOS, C. M.; SILVA, S. H. M. G. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 160, p. 11-23, 2017. <https://doi.org/10.1016/j.jastp.2017.04.001>.

SHEVADE, S. K. KEERTHI, S. S. BHATTACHARYYA, C. MURTHY, K. R. K. Improvements to the SMO Algorithm for SVM Regression. **IEEE TRANSACTIONS ON NEURAL NETWORKS**, v. 11, n. 5, p. 1188-1193, 2000. DOI: [10.1109/72.870050](https://doi.org/10.1109/72.870050).

VAPNIK, V. N. The nature of Statistical learning theory. Springer-Verlag, New York, 1995.

VOYANT, C.; NOTTON, G.; KALOGIROU, S.; NIVET, M-L.; PAOLI, C.; MOTTE, F.; FOUILLOY, A. Machine learning methods for solar radiation forecasting: A review. **Renewable Energy**, v. 105, p. 569-582, 2017. <https://doi.org/10.1016/j.renene.2016.12.095>.

WOLDEGIYORGIS, T. A.; BENTI, N. E.; CHAKA, N. E.; SEMIE, A. G.; JEMBERIE, A. A. Estimating solar radiation using artificial neural networks: A case

study of Fiche, Oroma, Ethiopia. **Cogent Engineering**, v. 10, n. 1, 2023. <https://doi.org/10.1080/23311916.2023.2220489>.

ZHOU, Y.; LIU, Y.; WANG, D.; LIU, X.; WANG, Y. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. **Energy Conversion and Management**, v. 235, n. 1, 2021. <https://doi.org/10.1016/j.enconman.2021.113960>.

5.0 CONSIDERAÇÃO FINAL E GERAL

De acordo com os resultados encontrados, não houve diferença expressiva no desempenho estatístico quando se compara os modelos empíricos em relação ao MLP e SVM com dados locais, para as combinações envolvendo insolação, temperatura do ar e combinações híbridas, pois essas variáveis analisadas, possuem forte relação linear com a Hg, podendo acontecer que modelos empíricos e técnicas de machine learning cheguem a resultados semelhantes. Porém, a necessidade de avaliar outras possibilidades, como abordado nos parágrafos abaixo.

Para compreender melhor o desempenho das diferentes metodologias empregadas na estimativa da Hg, é necessário avaliar outras variáveis de entrada como radiação difusa e principalmente os elementos que mais interferem na transmissividade atmosférica como os aerossóis.

Nas técnicas de ML, a necessidade de avaliar algoritmos híbridos que ajustam os hiperparâmetros de forma automatizada, pois pode melhorar significativamente o desempenho dessa metodologia.