

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS / FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL**

Samuel da Costa Alves Basílio

**Um framework AutoML para transferência de aprendizado aplicado à
predição de irradiação solar em Minas Gerais**

Juiz de Fora

2026

Samuel da Costa Alves Basílio

**Um framework AutoML para transferência de aprendizado aplicado à
predição de irradiação solar em Minas Gerais**

Tese apresentada ao Programa de pós-graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Modelagem Computacional. Área de concentração:

Orientador: Titulação Prof. DSC Leonardo Goliatt da Fonseca

Juiz de Fora

2026

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Basílio, Samuel da Costa Alves.

Um framework AutoML para transferência de aprendizado aplicado à
predição de irradiação solar em Minas Gerais / Samuel da Costa Alves
Basílio. – 2026.

113 f. : il.

Orientador: Prof. DSC Leonardo Goliatt da Fonseca

Tese (Doutorado) – Universidade Federal de Juiz de Fora, Instituto de
Ciências Exatas / Faculdade de Engenharia. Programa de pós-graduação
em Modelagem Computacional, 2026.

1. AutoML. 2. Transferência de Aprendizado. 3. Aprendizado de
Máquina. 4. Irradiação Solar. 5. Minas Gerais. I. Fonseca, Leonardo
Goliatt da. II. Título.

Samuel da Costa Alves Basílio

**Um framework AutoML para transferência de aprendizado aplicado à
predição de irradiação solar em Minas Gerais**

Tese apresentada ao Programa de pós-graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Modelagem Computacional. Área de concentração:

Samuel da Costa Alves Basílio

Um framework AutoML para transferência de aprendizado aplicado à predição de irradiação solar em Minas Gerais

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Modelagem Computacional. Área de concentração: Modelagem Computacional.

Aprovada em 06 de abril de 2026.

BANCA EXAMINADORA

Prof. Dr. Leonardo Goliatt da Fonseca - Orientador
Universidade Federal de Juiz de Fora

Prof^a. Dr^a. Érica da Costa Reis Carvalho
Universidade Federal de São João del-Rei

Prof. Dr. Fernando Rodrigues Trindade Ferreira
Universidade do Estado do Rio de Janeiro

Prof^a. Dr^a. Eliane Da Silva Christo
Universidade Federal de Juiz de Fora

Prof. Dr. Wanderlei Malaquias Pereira Junior
Universidade Federal de Catalão

Juiz de Fora, 22/04/2026.



Documento assinado eletronicamente por **Leonardo Goliatt da Fonseca, Professor(a)**, em 22/04/2026, às 17:37, 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 **Eliane da Silva Christo, Professor(a)**, em 22/04/2026, às 19:39, 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 **Erica da Costa Reis Carvalho, Usuário Externo**, em 23/04/2026, às 10:24, 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 **Wanderlei Malaquias Pereira Junior, Usuário Externo**, em 25/04/2026, às 11:35, 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 **Fernando Rodrigues Trindade Ferreira, Usuário Externo**, em 25/04/2026, às 13:51, 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 Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2955320** e o código CRC **475F39DF**.

AGRADECIMENTOS

Agradeço a Deus por sua misericórdia que me perdoa e graça que me salva.

À igreja que intercedeu por mim, eu sou devedor.

Aos meus pais pela herança incorruptível que me passaram.

À Du, por sua companhia e amor. Meu coração está nela confiado.

À Ana e à Emanuela, que são meu galardão.

Aos meus professores, especialmente o Leonardo Goliatt, pela orientação, ensino, correções, tempo, conselhos, etc. Sejam honrados por isso.

A todos que direta ou indiretamente me ajudaram, apoiaram e estiveram presentes.

Muito obrigado.

O Senhor Deus me deu uma língua erudita, para que eu saiba dizer boa palavra ao cansado. Ele me desperta todas as manhãs; desperta o meu ouvido para que eu ouça como aqueles que aprendem.

Isaías 50:4

RESUMO

Este trabalho apresenta o desenvolvimento de um framework de Automated Machine Learning (AutoML) integrado a técnicas de Transfer Learning para a predição da irradiação solar no estado de Minas Gerais, Brasil. A principal aplicação está na previsão de valores de irradiação em regiões com escassez de dados meteorológicos, permitindo previsões eficientes mesmo em locais desprovidos de estações de medição física por meio da transferência de conhecimento de regiões com dados abundantes. A metodologia utilizou um banco de dados composto por 134 estações meteorológicas, sendo 67 em Minas Gerais e 67 em estados vizinhos, abrangendo mais de 20 anos de medições em 129 cidades. O framework automatiza etapas críticas como a seleção de características, a otimização de hiperparâmetros e a escolha do modelo de aprendizado, destacando a eficiência do algoritmo Extreme Learning Machine (ELM), que apresentou desempenho superior a modelos como ANN, SVR, EN, MARS e XGBoost em testes preliminares. Foram exploradas diferentes abordagens, incluindo modelos pontuais para cada estação, modelos geolocalizados com acréscimo de dados vizinhos e o aprendizado por transferência, onde o treinamento é realizado sem dados do local alvo. Os resultados demonstraram que a adição de dados geolocalizados de cidades próximas para treinar modelos em pontos com estações existentes foi, em sua maioria, prejudicial ou irrelevante. Contudo, a aplicação de Transfer Learning para predição em locais sem estações mostrou-se viável e eficaz, com o desempenho aumentando conforme o número de cidades no treinamento é ampliado. A utilização de um algoritmo genético para a seleção automática das cidades que compõem a base de treino apresentou resultados substancialmente superiores à seleção baseada puramente em distância geográfica. O trabalho conclui que o framework proposto representa um avanço significativo para o planejamento de sistemas de energia fotovoltaica e gestão agrícola, provendo previsões precisas em locais com infraestrutura de monitoramento limitada.

Palavras-chave: autoML; aprendizado de máquina; irradiação solar; transfer learning; Minas Gerais.

ABSTRACT

This work presents the development of an Automated Machine Learning (AutoML) framework integrated with Transfer Learning techniques for the prediction of solar irradiation in the state of Minas Gerais, Brazil. The primary application lies in forecasting irradiation values in regions with a scarcity of meteorological data, enabling efficient predictions even in locations lacking physical measurement stations through the transfer of knowledge from regions with abundant data. The methodology utilized a database composed of 134 meteorological stations, with 67 in Minas Gerais and 67 in neighboring states, covering more than 20 years of measurements across 129 cities. The framework automates critical steps such as feature selection, hyperparameter optimization, and model selection, highlighting the efficiency of the Extreme Learning Machine (ELM) algorithm, which outperformed models such as ANN, SVR, EN, MARS, and XGBoost in preliminary tests. Different approaches were explored, including point-based models for each station, geolocated models with the addition of neighboring data, and transfer learning, where training is performed without data from the target site. The results demonstrated that the addition of geolocated data from nearby cities to train models at points with existing stations was, for the most part, detrimental or irrelevant. However, the application of transfer Learning for prediction in locations without stations proved to be viable and effective, with performance increasing as the number of cities in the training set expanded. The use of a genetic algorithm for the automatic selection of the cities that make up the training base yielded substantially superior results compared to selection based purely on geographic distance. The work concludes that the proposed framework represents a significant advancement for photovoltaic energy system planning and agricultural management, providing accurate predictions in locations with limited monitoring infrastructure.

Keywords: autoML; machine learning; solar irradiation; transfer learning; Minas Gerais.

LISTA DE ILUSTRAÇÕES

Matriz elétrica brasileira	16
Produção de energia no Brasil	16
Produção de energia solar no Brasil	17
Mapa das cidades estudadas	34
Plot das variáveis temporais da cidade de Juiz de Fora	36
Biomass dos locais estudados	38
Ilustração de uma ANN	42
Processo de otimização para cada modelo	48
Formação da base geolocalizada	49
Agrupamento de cidades vizinhas	50
Criação de banco de dados para o processo de transferência de aprendizado	51
Etapas de experimentos realizados	52
Diagramas de dispersão para as melhores execuções individuais de cada estação	57
Performance Profiles	58
<i>Heatmap</i> da utilização das cidades na formação da base de treino	59

LISTA DE TABELAS

Tabela 1 – Variáveis utilizadas	35
Tabela 2 – Métricas e indicadores de desempenho	40
Tabela 3 – Parametros do algoritmo de otimização.	52
Tabela 4 – Parametros utilizados para os modelos de aprendizado de máquina.	53
Tabela 5 – Áreas normalizadas dos perfis de desempenho - Etapa 1	54
Tabela 6 – Áreas normalizadas dos perfis de desempenho - Etapa 2	58
Tabela 7 – Média das métricas de desempenho para a cidade de Juiz de Fora	59
Tabela S1 – Estações meteorológicas automáticas	78
Tabela S2 – Resultados	92
Tabela S3 – Comparação com literatura	99
Tabela S4 – Utilização das cidades	105

LISTA DE SIGLAS E ABREVIATURAS

- R^2 Determination Coefficient. 28– 30, 40, 56, 57, 59
- AC** Ant Colony. 24
- ANN** Artificial Neural Network. 9, 13, 23, 26, 27, 30, 31, 41, 42, 53
- CNN** Convolutional Neural Network. 25
- DE** Differential Evolution. 24
- ELM** Extreme Learning Machine. 13, 23, 24, 31, 42, 43, 53
- EN** Elastic Net. 13, 43, 53
- GB** Gradient Boosting. 23
- INMET** Instituto Nacional de Meteorologia. 31
- LSTM** Long Short-Term Memory. 25
- MAE** Mean Absolute Error. 40, 59
- MAPE** Mean Absolute Percentage Error. 40
- MARS** Multivariate Adaptive Regression Spline. 13, 43, 53
- MBE** Mean Bias Error. 40, 59
- MLP** Multilayer Perceptron. 29
- ONU** Organização das Nações Unidas. 19
- PSO** Particle Swarm Optimization. 23, 24
- RF** Random Forest. 23
- RMSE** Root Mean Square Error. 28, 40, 47, 48, 59
- RR** Ridge Regression. 31
- SGA** Simple Genetic Algorithm. 13, 23, 45, 52
- SLFN** Single Hidden Layer Feedforward Neural Network. 42, 43
- SVM** Support Vector Machine. 23, 27, 30, 31

SVR Support Vector Regression. [13](#), [44](#), [45](#), [53](#)

WI Willmott Index of Agreement. [40](#), [59](#)

XGB Extreme Gradient Boosting. [13](#), [44](#), [53](#)

SUMÁRIO

1	INTRODUÇÃO	15
1.1	CONTEXTUALIZAÇÃO	15
1.2	OBJETIVOS DA PESQUISA	20
1.3	ORGANIZAÇÃO DO TEXTO	21
2	REVISÃO BIBLIOGRÁFICA	23
2.1	TRABALHOS DE REVISÃO BIBLIOGRÁFICA	23
2.2	MODELOS HÍBRIDOS DE OTIMIZAÇÃO	24
2.3	MODELOS ESPAÇO-TEMPORAIS	25
2.4	TRABALHOS DE MODELAGEM DA IRRADIAÇÃO SOLAR EM MINAS GERAIS	27
2.5	TRABALHOS DE MODELAGEM DA IRRADIAÇÃO SOLAR EM OUTROS ESTADOS DO BRASIL	29
2.6	OUTROS TRABALHOS RELACIONADOS	32
3	MATERIAIS E MÉTODOS	33
3.1	DADOS	33
3.1.1	BIOMAS	37
3.2	FERRAMENTAL TEÓRICO	38
3.2.1	MÉTRICAS DE DESEMPENHO	39
3.2.2	PERFIS DE DESEMPENHO	40
3.2.3	MODELOS DE APRENDIZADO DE MÁQUINA	41
3.2.3.1	Artificial Neural Network	41
3.2.3.2	Extreme Learning Machine	42
3.2.3.3	Elastic Net	43
3.2.3.4	Multivariate Adaptive Regression Spline	43
3.2.3.5	Extreme Gradient Boosting	44
3.2.3.6	Support Vector Regression	44
3.2.4	ALGORITMOS DE OTIMIZAÇÃO	45
3.2.4.1	Simple Genetic Algorithm	45
3.2.5	TRANSFERÊNCIA DE APRENDIZADO	46
3.2.6	SELEÇÃO DE CARACTERÍSTICAS AUTOMATIZADA	47
3.3	METODOLOGIA	47
3.3.1	ABORDAGEM COMPUTACIONAL HÍBRIDA	47
3.3.2	MODELO GEOLOCALIZADO	48
3.3.2.1	SELEÇÃO AUTOMÁTICA DAS BASES DE TREINO	51
4	EXPERIMENTOS COMPUTACIONAIS, ANÁLISE E DISCUS- SÃO	52
4.1	PRIMEIRA ETAPA DE EXPERIMENTOS	54

4.2	SEGUNDA ETAPA DE EXPERIMENTOS	54
4.2.1	DISCUSSÃO	56
4.3	TERCEIRA ETAPA DE EXPERIMENTOS	58
4.3.1	DISCUSSÃO	60
4.4	PONTOS FORTES E LIMITAÇÕES DO MODELO	60
5	CONCLUSÃO	62
5.1	POSSÍVEIS DESDOBRAMENTOS	64
	REFERÊNCIAS	65
.1	APÊNDICE – Tabelas de dados e resultados	78
	ANEXO A – Contribuições à literatura	112

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

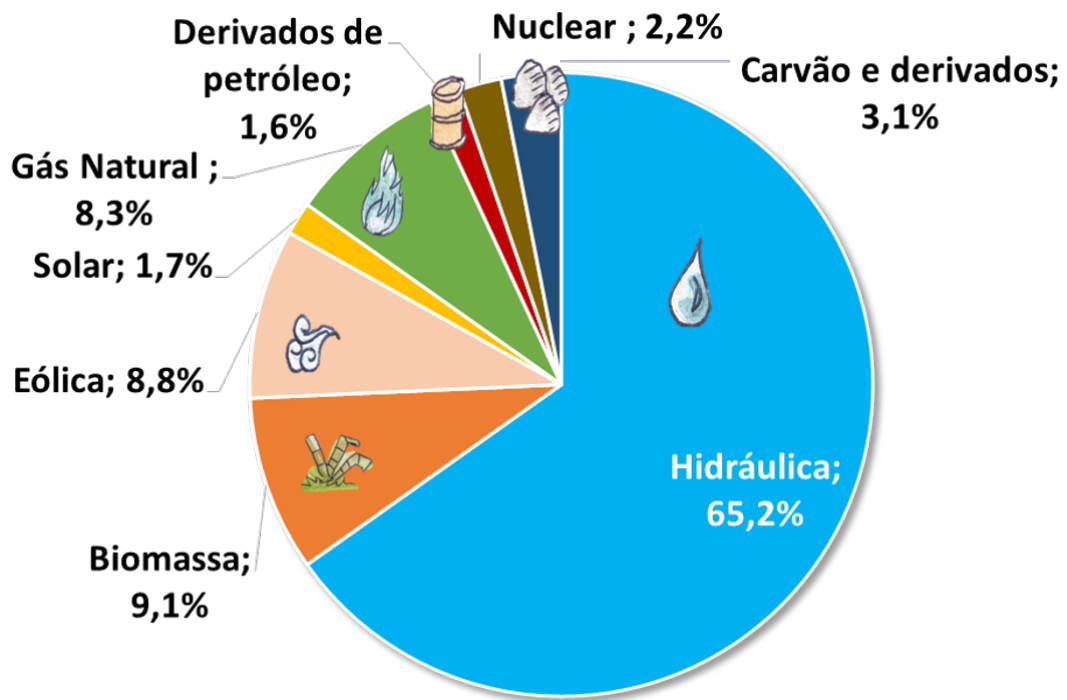
A previsão dos valores de irradiação solar é útil para diferentes áreas do conhecimento. Os principais modelos para cálculo da evapotranspiração necessitam desta medida (48, 49, 83) e, conseqüentemente modelos de crescimento de culturas (27, 94). A predição de outras variáveis meteorológicas também utiliza este dado(69, 32), assim como modelos de produção de energia (92, 41, 68). A eficiência dessa previsão é essencial tanto no planejamento de sistemas quanto na execução de ações de curto, médio e longo prazo nessas áreas do conhecimento (31). Além desta natural necessidade em prever eficientemente os valores de irradiação solar em diferentes áreas do conhecimento, nos últimos anos, o discurso da luta contra o aquecimento global e do desenvolvimento de energia renovável tornou-se o objetivo dos esforços conjuntos da maioria dos países (17). Juntamente com a energia eólica e hídrica, a energia fotovoltaica é um dos tipos de energia renovável que temos à disposição. No entanto, a natureza da intermitência da energia fotovoltaica e a incerteza relacionada às previsões são problemas difíceis que devem ser superados para manter a estabilidade do sistema de energia. Considerando que a medição física da irradiação solar, devido ao custo e à tecnologia envolvida, não é possível para todos os locais do mundo, a previsão de valores futuros geralmente se restringe a áreas com estações de medição (67).

No Brasil, a principal fonte de energia elétrica provém dos recursos hídricos, com potencial inventariado de 176 GW. Contudo, estima-se que 108 GW já foram aproveitados e, sobre o restante, mais da metade do potencial está localizada nas regiões hidrográficas Amazônica e Tocantins-Araguaia, onde há grandes extensões de áreas protegidas por unidades de conservação e terras indígenas. Desta forma, a expansão da utilização dos recursos hidrelétricos tem enfrentado alguns desafios recentes, a saber, os impactos socioambientais negativos causados na implantação de projetos hidrelétricos de grande porte, os custos elevados de investimentos nos anos iniciais do projeto e a dificuldade de desenvolver estações geradoras próximas aos grandes centros de consumo, o que resulta na necessidade de investimentos adicionais em longas linhas de transmissão para escoamento da produção de eletricidade (79).

A Figura 1 mostra a distribuição entre as fontes geradoras de energia elétrica no Brasil. Apesar de ecologicamente saudável, a extrema dependência dos recursos hídricos no Brasil tem sido um problema, pois a escassez hídrica gera imediatamente um aumento na utilização de fontes poluentes e no custo da energia. Desta forma, muito tem sido investido na diversificação da matriz energética do Brasil. A Figura 2 mostra o crescimento da produção energética (não apenas a elétrica) ao longo dos anos. Nesta figura notamos como a geração de energia fotovoltaica ainda é incipiente e, ainda que o Brasil possua uma capacidade gigantesca de geração de energia através da irradiação solar, este recurso ainda

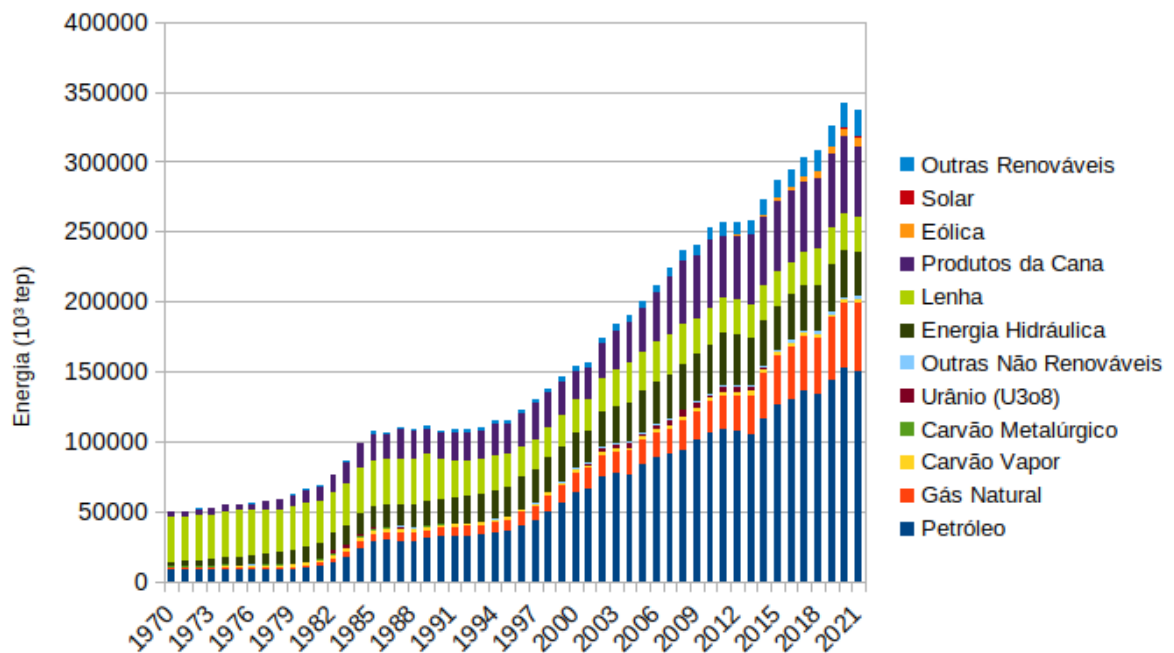
não é devidamente aproveitado (80).

Figura 1 – Matriz elétrica brasileira



Fonte: Ministério de Minas e Energia (80)

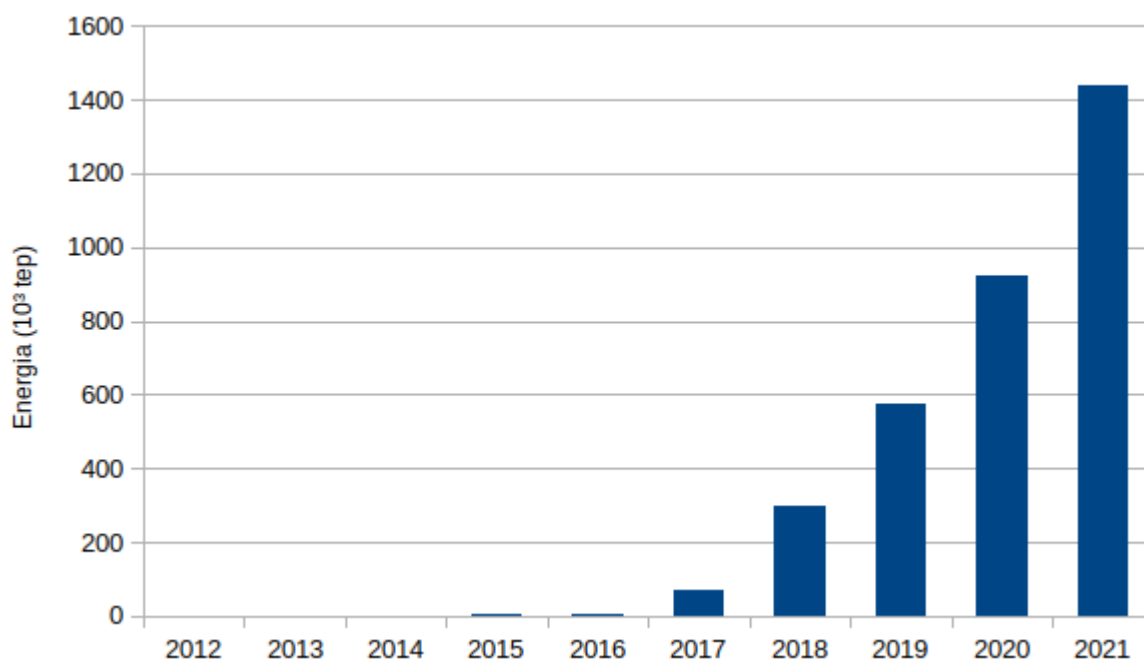
Figura 2 – Produção de energia no Brasil



Nota: Valores em toneladas equivalentes de petróleo.

Atualmente a maior parte da expansão da produção de energia fotovoltaica do Brasil está na geração distribuída, que vem aumentando de forma contínua nos últimos anos, conforme podemos observar na Figura 3. Comparando somente o crescimento da micro e minigeração distribuída, de 2020 para 2021, a geração solar teve um aumento de 84%, representando 88.3% do total, sendo a principal fonte responsável pelo aumento registrado na micro e minigeração distribuída (80).

Figura 3 – Produção de energia solar no Brasil



Nota: Valores em toneladas equivalentes de petróleo.

Existe atualmente no Brasil uma expectativa de mudança na forma que o setor elétrico realiza a venda de energia elétrica. Pelo menos dois projetos de lei estão em análise e tem como principal objetivo a abertura do mercado de distribuição de energia, hoje disponível apenas aos os grandes consumidores, como indústrias e shoppings, também aos consumidores residenciais (20, 22, 21). Em países onde o mercado livre já está completamente implantado, estudos apontam como a acurácia nas previsões impacta financeiramente consumidores e distribuidores (4, 107, 89).

Quando a previsão dos valores de irradiação solar é necessária ou desejada, e não há estações meteorológicas, modelos de previsão são utilizados (7, 50). Modelos empíricos têm sido utilizados para esse propósito há muito tempo. No entanto, estes geralmente apresentam fortes generalizações e não fornecem consistentemente os melhores resultados(70). Recentemente, o progresso nas técnicas de aprendizado de máquina permitiu que recursos computacionais substituíssem, de forma parcial ou integral, os modelos empíricos tradicionais, alcançando resultados promissores(84, 58). Contudo, para

garantir a eficiência na execução, é imprescindível que o método de classificação escolhido seja devidamente calibrado.

Existem muitos desafios na previsão da irradiação solar; portanto, diferentes técnicas são empregadas em vários locais (38, 42). Além disso, a literatura mostra claramente que um modelo de previsão que funciona bem em um local não é necessariamente o melhor em outro, assim como as melhores variáveis para um local não são necessariamente as melhores para outro (60). Ao selecionar o modelo de aprendizado de máquina mais adequado, é essencial considerar se os dados são originários de estações meteorológicas ou satélites (55) e a granularidade de previsão desejada. (113, 37, 57, 114). Independentemente do tipo de dado e modelo, métodos híbridos que otimizam os hiperparâmetros dos modelos de aprendizagem apresentam melhores resultados (109, 3, 66).

No Brasil, a principal fonte de produção da rede elétrica são os recursos hidrelétricos. No entanto, a exploração desses recursos está se aproximando do seu limite, e a forte dependência de uma fonte de energia sazonal causa distúrbios e flutuações de preços durante os períodos de seca. Dessa forma, investimentos consideráveis têm sido feitos na produção de energia fotovoltaica, principalmente em micro e minigeração distribuída.

A geração distribuída (GD) de energia elétrica a partir de fonte solar fotovoltaica tem se destacado no Brasil como uma alternativa sustentável e descentralizada para o fornecimento de energia. Segundo dados da ANEEL e da ABSOLAR(2, 5), 99,26% da potência instalada em microgeração (até 75 kW) e minigeração (acima de 75 kW até 5 MW) distribuída no país é proveniente de fonte solar fotovoltaica, demonstrando a ampla predominância dessa tecnologia no setor. Atualmente, o Brasil conta com mais de 3,3 milhões de sistemas solares conectados à rede, beneficiando aproximadamente 4,96 milhões de unidades consumidoras por meio do Sistema Integrado de Compensação de Energia Elétrica (SIN).

Minas Gerais é, em termos econômicos e populacionais, o terceiro maior estado do Brasil, com um território de 586.852,35 km² e uma população de 20 milhões de habitantes. Seu Produto Interno Bruto (PIB) é o terceiro maior do Brasil e, com uma demanda mensal de energia de aproximadamente 5 vezes 106 MWh, o estado está em constante processo de expansão de sua matriz energética. Minas Gerais também é o segundo estado brasileiro com maior capacidade instalada em geração solar distribuída, totalizando 4.609,5 MW, representando 12,4% da capacidade nacional. Isso demonstra a liderança de Minas Gerais no cenário de geração distribuída, resultado de políticas públicas estaduais, atratividade de investimentos e amplo engajamento público na transição energética. No estado de Minas Gerais, existe uma quantidade considerável de estações meteorológicas mantidas por diferentes instituições (81). Contudo, parte das que tem aparelhagem capaz de medir a irradiação solar global é antiga e não transmitem seus dados de forma online, muitas vezes, apresentando lacunas nas séries temporais de suas medições (104, 33). Mesmo que estas

estações sejam suficientes para previsões específicas e instantâneas, como temperatura e umidade, muitos problemas se beneficiam ou até mesmo requerem previsões com prazos mais estendidos (115). Juiz de Fora, localizada na Zona da Mata mineira, é a décima maior cidade produtora, com 39,3 MW instalados. É uma das maiores cidades da região em termos econômicos e populacionais. Possui um centro meteorológico do INMET(65) em operação desde 2007, que capta diferentes informações meteorológicas, que serão detalhadas posteriormente. Dessa forma, Minas Gerais é um dos principais polos nacionais de geração solar distribuída, contribuindo significativamente para a expansão das energias renováveis no Brasil.

Além disso, o setor de produção de energia solar fotovoltaica no Brasil conta com 153 fabricantes de kits fotovoltaicos cadastrados no FINAME do BNDES e outros fabricantes de inversores, rastreadores solares, módulos, baterias e string boxes. Em termos de perfil de consumo, o segmento residencial responde pela maior parte da potência instalada (49,45%), seguido pelos setores comercial e de serviços (28,45%) e rural (13,63%), refletindo o protagonismo de consumidores individuais e pequenas empresas na adoção da tecnologia.

Outro ponto que deve ser elucidado é a participação do Brasil como signatário dos Objetivos de Desenvolvimento Sustentável da [Organização das Nações Unidas \(ONU\)](#)(85). Este projeto apresenta 17 pontos que buscam mundialmente acabar com a pobreza, proteger o meio ambiente e o clima e garantir que as pessoas, em todos os lugares, possam desfrutar de paz e de prosperidade. O trabalho aqui proposto está intimamente relacionado com o 7º objetivo denominado Energia limpa e acessível. Neste item, transcrevemos os objetivos específicos onde a presente proposta pode contribuir de forma efetiva, são eles:

1. Até 2030, assegurar o acesso universal, confiável, moderno e a preços acessíveis a serviços de energia;
2. Até 2030, aumentar substancialmente a participação de energias renováveis na matriz energética global;
3. Até 2030, dobrar a taxa global de melhoria da eficiência energética;
4. Até 2030, reforçar a cooperação internacional para facilitar o acesso a pesquisa e tecnologias de energia limpa, incluindo energias renováveis, eficiência energética e tecnologias de combustíveis fósseis avançadas e mais limpas, e promover o investimento em infraestrutura de energia e em tecnologias de energia limpa;
5. Até 2030, expandir a infraestrutura e modernizar a tecnologia para o fornecimento de serviços de energia modernos e sustentáveis para todos nos países em desenvolvimento, particularmente nos países menos desenvolvidos, nos pequenos Estados insulares em

desenvolvimento e nos países em desenvolvimento sem litoral, de acordo com seus respectivos programas de apoio.

Considerando todos os pontos elencados acima, fica clara a utilidade de um modelo eficiente para a previsão dos valores de irradiação solar em Minas Gerais, seja pela expectativa do aumento na produção de energia fotovoltaica, pelo estado atual da produção, especialmente a distribuída, ou pela utilização do conhecimento de valores futuros de irradiação solar em outras áreas como agricultura e meteorologia. Neste trabalho, apresentamos um framework para otimizar modelos computacionais de predição de irradiação solar aplicados a dados do estado de Minas Gerais, Brasil.

1.2 OBJETIVOS DA PESQUISA

Neste trabalho, apresentamos um modelo computacional para todo o estado de Minas Gerais e a metodologia para seu desenvolvimento. Este modelo utiliza dados de 67 estações meteorológicas distribuídas por todas as regiões do estado, além de dados de outras 67 estações em estados vizinhos, com medições realizadas ao longo de mais de 20 anos em 129 cidades. Diferentes abordagens são estudadas, manipulando os dados utilizados nos bancos de dados de treinamento e teste, de modo a responder às seguintes questões:

- A adição de dados de geolocalização para a predição da irradiação solar melhora os resultados?
- É possível prever eficientemente os valores de irradiação solar em um local sem estações meteorológicas?
- Como identificar o melhor conjunto de dados para treinamento de um modelo de aprendizado?

Três abordagens distintas são comparadas neste trabalho. Em todas utilizamos um algoritmo otimizador para selecionar o modelo de aprendizado de máquina mais eficiente, os melhores hiperparâmetros e variáveis. Na primeira abordagem, adicionamos dados de estações vizinhas aos dados de treinamento de cada estação, com o intuito de verificar como a adição de dados de geolocalização influencia os resultados. Em um segundo momento, removemos do banco de dados de treinamento os dados do local onde a predição será realizada, mantendo apenas os dados das estações vizinhas. Dessa forma, obtemos um banco de dados de treinamento sem dados do local onde a predição será feita. Os dois métodos são comparados com a execução pontual de cada estação. A terceira abordagem é semelhante à segunda, com a melhoria de que as cidades selecionadas para comporem a base de treino dos modelos de aprendizado também são selecionadas pelo algoritmo otimizador.

Sendo assim, temos como objetivo geral deste trabalho **Desenvolver um framework de Automated Machine Learning (AutoML) integrado a técnicas de transferência de aprendizado para a predição da irradiação solar no estado de Minas Gerais, Brasil, visando suprir a carência de dados em regiões sem estações de medição física.** Como objetivos específicos temos:

- Implementar uma abordagem computacional híbrida que automatize a seleção de modelos, a otimização de hiperparâmetros, a seleção de características e a escolha da base de treino para predição de irradiação solar;
- Avaliar o impacto da adição de variáveis geoespaciais e dados de estações vizinhas no desempenho de modelos pontuais já existentes;
- Validar a eficácia da transferência de aprendizado ao treinar modelos exclusivamente com dados de localidades distintas para predição em locais sem uma base de dados histórica;
- Otimizar a seleção das bases de treinamento por meio de algoritmos otimizadores, superando critérios puramente geográficos na escolha das cidades que compõem o banco de dados de treino.

1.3 ORGANIZAÇÃO DO TEXTO

Esta tese está estruturada em cinco capítulos principais, organizados de forma a apresentar desde a fundamentação teórica até os resultados obtidos com o framework proposto:

Capítulo 1 – Introdução: Apresenta a contextualização do problema, destacando a importância da predição da irradiação solar no cenário energético e agrícola de Minas Gerais, além de definir os objetivos e as questões de pesquisa que norteiam este trabalho;

Capítulo 2 – Revisão Bibliográfica: Realiza um levantamento do estado da arte, abordando trabalhos de revisão sistemática, modelos híbridos de otimização, modelos espaço-temporais e estudos específicos sobre a modelagem da irradiação solar em Minas Gerais e em outros estados do Brasil;

Capítulo 3 – Materiais e Métodos: Detalha o banco de dados utilizado, composto por 134 estações meteorológicas. Descreve o ferramental teórico, incluindo as métricas de desempenho, os perfis de desempenho e os modelos de aprendizado de máquina testados. Explica ainda a metodologia do framework AutoML, o uso de algoritmos genéticos para seleção de características e a estratégia de transferência de aprendizado (*Transfer Learning*).

Capítulo 4 – Experimentos Computacionais, Análise e Discussão: Apresenta os resultados das três etapas de experimentos. A primeira etapa justifica a escolha do modelo

ELM. A segunda analisa os cinco cenários propostos para validação da geolocalização e transferência de aprendizado básica. A terceira etapa foca na seleção automática das bases de treino para a cidade-alvo de Juiz de Fora, discutindo a eficácia do otimizador na escolha das cidades componentes.

Capítulo 5 – Conclusão: Sintetiza as principais conquistas do trabalho, confirma a viabilidade da predição em locais sem estações meteorológicas e aponta limitações e possíveis desdobramentos para pesquisas futuras, como a automação do fluxo de dados online.

Por fim, o trabalho inclui Apêndices com tabelas detalhadas das estações e dos resultados numéricos de todos os cenários, além de um Anexo que compila as contribuições deste estudo para a literatura acadêmica.

2 REVISÃO BIBLIOGRÁFICA

2.1 TRABALHOS DE REVISÃO BIBLIOGRÁFICA

Alguns trabalhos se propõem a resumir o estado da arte em determinadas áreas. Abaixo listamos alguns que mostram o que foi realizado na área de previsão de irradiação solar durante os últimos anos.

Yadav and Chandel (109) faz um estudo especificamente sobre ANN aplicados à problemas de predição de irradiação solar. Entre outros pontos, os autores concluem que as variáveis espaciais como latitude e longitude devem ser utilizadas para que previsões possam ser feitas em locais onde não existem estações meteorológicas e que os modelos devem ser treinados com técnicas de otimização como Particle Swarm Optimization (PSO) e Simple Genetic Algorithm (SGA).

Zhang et al. (113) faz uma revisão bibliográfica de modelos de previsão de irradiação solar e os compara considerando as métricas utilizadas, o tipo de modelo (empírico ou de aprendizado de máquina), as variáveis utilizadas e a granularidade temporal de cada artigo.

Das et al. (37) classifica detalhadamente os tipos de trabalhos considerando o horizonte e a técnica de previsão. O trabalho também ilustra a correlação entre as variáveis. Os autores concluem que os modelos ANN e SVM são os que melhor performam com um tempo de execução aceitável. O desenvolvimento de modelos híbridos de otimização e que utilizam técnicas de *ensemble* também é incentivado.

Akhter et al. (3) foca na importancia na utilização de metaheurísticas, no sentido de métodos que utilizam mais de um modelo de aprendizado de máquina para composição do resultado final, métodos que otimizam os hiperparâmetros dos modelos, que combinam modelos de aprendizado de máquina com modelos estatísticos, etc.

Guermoui et al. (57) apresenta um estudo comparativo entre diferentes modelos híbridos, classificando-os como abordagens de *ensemble*, abordagem em *clusterização* dos dados, abordagens de baseadas na decomposição do sinal de entrada, abordagens que consideram os componentes residuais e modelos híbridos de otimização, sendo este último o mais detalhado.

de Freitas Viscondi and Alves-Souza (38) faz uma revisão bibliográfica sobre *Big Data* aplicado à predição de geração elétrica fotovoltaica. Apesar de, se compararmos à outros trabalhos de revisão bibliográfica, a metodologia de revisão sistemática deste trabalho ter retornado poucos artigos, podemos destacar neste trabalho a contagem dos modelos mais utilizados: SVM, ANN, ELM, GB e RF, nesta ordem.

Zhou et al. (114) Faz uma revisão bibliográfica de modelos de aprendizado da máquina aplicados à previsão de irradiação solar. Os autores categorizam os tipos de

modelos utilizados, as variáveis de entrada, suas formas de obtenção e os métodos para seleção destas variáveis.

Carneiro et al. (24) apresenta algumas métricas interessantes sobre a distribuição dos trabalhos que tratam sobre previsão de irradiação solar. Cerca de 42% dos artigos analisados desenvolveram abordagens híbridas, 83% realizaram previsão de curto prazo, e mais de 78% tiveram, como previsão objetivo, energia fotovoltaica, radiação solar e irradiação solar. Considerando a previsão espacial, 66% trabalham com um único ponto.

Wu et al. (108) classifica detalhadamente os métodos utilizados para previsão de energia fotovoltaica. Entre as conclusões do trabalho, os autores destacam a importância de avaliar novas técnicas de aprendizado de máquina na previsão de irradiação solar e de utilizar algoritmos de otimização juntamente com os modelos de aprendizado de máquina para alcançar melhores resultados.

2.2 MODELOS HÍBRIDOS DE OTIMIZAÇÃO

Diferentes trabalhos se classificam como apresentando ou utilizando modelos híbridos (109, 57). Alguns utilizam técnicas para escolha de diferentes modelos de aprendizado de máquina, outros efetivamente misturam dois modelos utilizando técnicas de *ensemble*. Nesta seção abordaremos alguns trabalhos que utilizam algoritmos de otimização para seleção do melhor conjunto de hiperparâmetros de um modelo de aprendizado de máquina. Em todo o texto esta é a nossa definição de modelo híbrido de otimização.

Ghimire et al. (56) propõe um modelo híbrido de otimização com ELM e DE utilizando AC como seletor de características para previsão de irradiação solar global. Este modelo e algumas de suas variações são analisados observando diferentes granularidades temporais, mas focando na média mensal. Os dados são provenientes de duas bases de dados, uma proveniente de medições terrestres, com um total de 87 diferentes variáveis, outra com dados de medições por satélite, com 67 diferentes variáveis, observados entre 2002 até 2017, em duas cidades da Austrália. O trabalho também apresenta uma boa revisão bibliográfica, categorizando os trabalhos em modelos empíricos, modelos de aprendizado de máquina e modelos híbridos.

Perera et al. (88) utiliza um PSO como técnica de *ensemble* na formação de um modelo a partir de 5 diferentes outros métodos de previsão. O estudo mostra como o algoritmo de otimização pode ser utilizado tanto para geração de um novo modelo a partir da combinação de métodos base, assim como para otimização dos hiperparâmetros dos métodos base. A base de dados provém de 25 residências produtoras de energia fotovoltaica nos Estados Unidos.

2.3 MODELOS ESPAÇO-TEMPORAIS

Consideramos como modelos espaço-temporais aqueles que de alguma forma consideram a geolocalização dos dados na sua metodologia. Muitos buscam agrupar os dados geograficamente distribuídos em *Grids* que podem facilmente ser utilizados com modelos [Convolutional Neural Network \(CNN\)](#) e [Long Short-Term Memory \(LSTM\)](#). Alguns destes trabalhos também podem ser classificados como modelos híbridos de otimização e/ou que utilizam técnicas de *ensemble*.

Liu et al. (76) utiliza dados de satélite para criar um *Grid* com os valores das variáveis analisadas. Este *grid* representa geograficamente os valores obtidos e é aplicada a uma [CNN](#) para que as previsões em pontos específicos possam ser realizadas. O trabalho é continuado em Liu et al. (75), onde os autores acrescentam a análise da incerteza probabilística para melhorarem os resultados.

Chai et al. (28) utiliza uma base de dados sintética para criar uma série de 56 estações distribuídas no estado do Arizona, Estados Unidos. Como a posição das estações é arbitrariamente escolhida, os autores facilmente transformam toda a base em um *Grid* e então aplicam juntamente [CNN](#) e [LSTM](#). Como os dados são gerados sinteticamente há uma preocupação dos autores em inserir e analisar diverentes níveis de ruído na base de dados.

Em Kumari and Toshniwal (71) dados de 23 estações na Califórnia, Estados Unidos, são analisados através de um modelo híbrido [CNN- LSTM](#), otimizados através de grid search, onde a ordem da análise dos dados espaciais e temporais é estudada de forma mais detalhada. Por fim os autores apresentam os resultados observando o comportamento dos modelos de forma anual, dividindo entre as quatro estações e sob três condições de céu.

Apesar da metodologia dos trabalhos acima ser bem diferente da proposta desta tese, a começar pela forma de captura dos dados, este tipo de trabalho pode ser, no futuro, um interessante conjunto de *benchmark*, visto que os dados utilizados são públicos e mundiais. Desta forma, seria possível comparar as medições feitas por estações e satélites, ou mesmo combinar o resultado destas diferentes abordagens para geração de resultados mais precisos.

Em Yang et al. Yang et al. (111) dados de 10 estações em Singapura são analisados. O objetivo do trabalho é apresentar uma metodologia de modelos numéricos para remapear a distância entre as estações. Após o remapeamento os dados são tratados com modelos estatísticos. O interessante deste trabalho é a aplicação de modelos *kriging* para a construção de mapas de contorno.

Yang et al. (110) propõe e avalia o desempenho de um modelo com horizonte de previsão de um minuto até duas horas à frente, usando dados de 13 locais na Califórnia e 5 locais no Colorado, Estados Unidos. O modelo proposto foca em uma previsão em

tempo real com um intervalo móvel de observação. Para tal, o modelo considera os valores históricos de irradiação solar no local, em locais próximos, o horizonte de medição, a diferença de tempo entre a medição do local alvo com os locais próximos e um índice para equilibrar a relação entre custo computacional e precisão do modelo. Uma interessante estratégia apresentada, que também pode ser útil no presente trabalho, principalmente na escolha do conjunto de estações vizinhas a serem consideradas, é a utilização da correlação dos valores de irradiação solar da estação alvo com as estações vizinhas.

Gutierrez-Corea et al. (61) utiliza ANN para predição de irradiação solar em pequenos horizontes de previsão. A hipótese, comprovada pelos autores, é que no conjunto de dados analisado o acoplamento dos dados de estações vizinhas como parâmetros de entrada para o modelo, juntamente com os dados das localizações das estações melhora o desempenho do modelo. Apesar de funcionar, esta abordagem cria modelos com muitos parâmetros, a saber, mais de 900 com os dados utilizados. É curioso que o limite de parâmetros do modelo aparece pela limitação do banco de dados utilizado, que permite um máximo de 1000 colunas. Desta forma, os autores limitam a observação às 10 estações mais próximas, cada uma com 9 variáveis.

Tascikaraoglu et al. (101) utiliza um modelo autorregressivo e mostra que nem todas as estações vizinhas necessariamente tem uma correlação significativa com a estação analisada. Apesar da premissa ser interessante, a base de dados continha apenas 5 estações.

Lan et al. (72) faz uma interessante aplicação prevendo a irradiação solar de um dia a frente para rotas marítimas. Uma estratégia interessante neste trabalho é a divisão da frequência da irradiação solar, onde as baixas frequências representam a variação diária causada pelo movimento periódico do sol e da terra, enquanto as altas frequências correspondem a distúrbios causados por nuvens em movimento. O tratamento espacial neste trabalho acontece através da escolha das estações mais próximas do navio. Nas simulações apresentadas um navio passa por 9 estações na costa oeste da China e durante o trajeto as quatro estações mais próximas são consideradas.

Por fim, Rodríguez et al. (92) desenvolve um modelo espaço-temporal para predições de curtíssimo prazo. Inicialmente o modelo é utilizado na previsão de irradiação solar e, num segundo momento, estes dados são utilizados no cálculo da potência de geradores fotovoltaicos. Dados de várias estações na Espanha, entre 2015 e 2017, foram utilizados. O modelo apresenta bons resultados em dias ensolarados, nublados e parcialmente nublados. Apesar do artigo mostrar uma grande quantidade de estações disponíveis, não fica claro se estas estações estão na mesma cidade e aparentemente toda a análise é feita somente na estação da cidade de Vitoria-Gasteiz, na Espanha.

2.4 TRABALHOS DE MODELAGEM DA IRRADIAÇÃO SOLAR EM MINAS GERAIS

Muitos trabalhos foram realizados durante os últimos anos buscando modelar da melhor forma a irradiação solar global, em Minas Gerais, no Brasil e no Mundo. No geral, observamos que tradicionalmente, no Brasil, os trabalhos utilizam de modelos empíricos para predição de valores futuros ou faltantes. Tais modelos são simples de serem utilizados e por vezes apresentam resultados suficientemente satisfatórios, contudo, esse padrão não é seguido em outros países, onde podemos encontrar uma vastidão de trabalhos que utilizam modelos computacionais para o mesmo fim e, com resultados bem mais interessantes.

Outra característica que podemos destacar, não somente nos trabalhos brasileiros, é que em muitos trabalhos, mesmo que várias estações sejam analisadas os resultados apresentados são apenas relatos pontuais unificados de várias estações em locais geograficamente convenientes (24). Trabalhos que efetivamente levam em consideração variáveis espaciais são mais escassos e focados em previsões de curtíssimo prazo (15). Muitas vezes os dados utilizados neste tipo de análise são provenientes de redes de sensores ou medições por satélite, e não de várias estações meteorológicas (46, 112).

Quando analisamos modelos computacionais, temos alguns estágios de desenvolvimento que costumam refletir na qualidade dos resultados encontrados. Os trabalhos mais antigos ou mesmo os que visivelmente são executados por pesquisadores sem expertise na computação, apenas aplicam modelos computacionais, em geral [Artificial Neural Network \(ANN\)](#) ou [Support Vector Machine \(SVM\)](#), sem se preocupar com os parâmetros ou testando diferentes combinações manualmente, sem nenhuma sistematização. Alguns trabalhos já se preocupam em buscar um conjunto ótimo dos hiperparâmetros e/ou o melhor modelo para determinado conjunto de dados. Por fim, encontramos trabalhos que buscam mesclar diferentes modelos, aproveitando as melhores características de cada abordagem e tentando suprimir suas deficiências. Estes costumam apresentar os melhores resultados.

Apesar dessa crescente demanda, poucos estudos publicados utilizam modelos computacionais para prever valores futuros de irradiação solar em Minas Gerais. Encontramos 17 trabalhos que abordam esse tema na região, e apenas 2 deles cobrem todo o estado, utilizando granularidade diária, um com modelos empíricos e o outro com modelos computacionais. A maioria dos artigos utiliza modelos empíricos para regiões específicas do estado (36, 8, 51, 25, 86, 96, 97, 91, 81). Alguns utilizam recursos computacionais para resolver modelos numéricos mais detalhados ou aplicam modelos empíricos como entrada para modelos computacionais (82, 78). As obras mais abrangentes são (33, 13).

Dantas et al. (36) é o trabalho mais antigo encontrado e apresenta um estudo para determinar as constantes do modelo linear de Angstrom para predição de irradiação solar utilizando dados da cidade de Lavras em Minas Gerais, com dados de dezembro de 2001 a novembro de 2002.

Barbosa et al. (8), Finzi et al. (51), Carvalho et al. (25) e Paula et al. (86) desmembraram um trabalho maior, publicado na forma de vários artigos no mesmo congresso, analisando em cada artigo uma região específica de Minas Gerais. Os artigos mostram a comparação dos resultados obtidos de 6 diferentes modelos empíricos com dados de 2008 até 2010. Em (8) 7 cidades são analisadas, em (51) 4, em (25) 9 e em (86) 6. Parte destas publicações foram republicadas com algumas modificações em Silva et al. (96) e Silva et al. (97). Em (96) os autores utilizam 9 modelos empíricos para estimar irradiação solar em 3 das 12 mesorregiões de Minas Gerais. Dados de 11 estações, nestas regiões foram utilizados com observações entre 2008 a 2010. Os melhores valores apresentados são de $RMSE=3.1$ e $R^2=0.7336$. Já em (97) dados de 4 estações em Minas Gerais, observados entre 2008 e 2010, aplicando 9 modelos empíricos para a previsão de irradiação solar global. Semelhantemente ao trabalho supracitado, este também foca mais na apresentação dos coeficientes de calibração encontrados do que nos resultados das previsões, apresentando apenas valores médios das métricas de cada modelo. Os valores médios de $RMSE = 3.1$ e $R^2=0.62$ são citados. Todos estes trabalhos apresentam os resultados agrupados pelos modelos utilizados e não pelas estações, impossibilitando uma comparação mais precisa com a presente proposta. Além disso, alguns resultados são apresentados apenas como gráficos, dificultando a averiguação dos resultados exatos.

Moraes and Miranda (82) utiliza o modelo do *European Center for Medium-Range Weather Forecast* para a predição de valores de precipitação, temperatura média, máxima e mínima do ar, irradiação solar e evapotranspiração de referência, observando 12 estações em Minas Gerais com dados de 2007. O trabalho apresenta bons resultados nas suas métricas, contudo, o horizonte de previsão é decendial e, por isso, não foi comparado com os resultados da presente proposta.

Tiba et al. (104) mapeia dados sobre irradiação solar e índices pluviométricos em Minas Gerais provenientes de diferentes bases mostrando como fatores geográficos influenciam estas medições. O trabalho apresenta um interessante conjunto de mapas mostrando, além dos fatores climáticos, objetivo do estudo, a localização e tipo de diversas estações meteorológicas, assim como seus mantenedores. Contudo o trabalho é apenas um mapeamento, e nenhum tipo de predição é feita.

Ramos et al. (91) avalia o desempenho de dois métodos empíricos de estimativa da irradiação global diária para diferentes localidades brasileiras. Foram analisadas 32 localidades distribuídas em diferentes regiões do Brasil, em um período variável para cada localidade que compreende entre os anos de 1999 a 2016. Apesar do foco do trabalho ser o Brasil como um todo, existem nas estações analisadas 7 em Minas Gerais, sendo que 2 delas também são avaliadas na presente proposta.

Monteiro and Martins (81) utiliza dados de 10 estações em Minas Gerais, observados entre 1999 e 2017, aplicando 13 modelos empíricos para a previsão de irradiação solar

global. Mesmo que o trabalho analise uma quantidade considerável de diferentes modelos, nenhum deles alcançou bons resultados.

(33) compilou um banco de dados com 51 estações, mostrando os desafios na coleta e limpeza dos dados. Quinze modelos empíricos diferentes foram aplicados a todas as cidades disponíveis. O melhor resultado encontrado pela métrica R^2 foi de 59,58%, com uma média de 36,97%. (13) utiliza a mesma base de dados do trabalho anterior, diferenciando-se pela aplicação de modelos computacionais, especificamente **Artificial Neural Networks** (ANN) and **Multivariate Adaptive Regression Spline** (MARS). A presente proposta utiliza estes trabalhos como principal parâmetro para comparação de resultados.

Outros dois trabalhos que não tratam diretamente da predição da irradiação solar, porém, estão fortemente relacionados ao presente artigo são (95), onde os autores utilizam uma abordagem computacional sobre o mesmo banco de dados, observando todo o estado de Minas Gerais por meio de rede neural artificial, floresta aleatória, máquina de vetores de suporte e regressão linear múltipla para a predição da evapotranspiração de referência, e (104), que mapeia dados de irradiação solar e precipitação em Minas Gerais a partir de diferentes bases, mostrando como fatores geográficos influenciam essas medições. O artigo apresenta um interessante conjunto de mapas que mostram, além dos fatores climáticos, o objetivo do estudo, a localização e o tipo de diversas estações meteorológicas e seus respectivos responsáveis. Contudo, o trabalho se limita ao mapeamento, não apresentando nenhuma predição.

Mais recentemente, em (54), os autores apresentaram um interessante trabalho, também utilizando modelos de aprendizado de máquina, para prever os valores de irradiação solar mensal em Minas Gerais. Os principais diferenciais entre o trabalho supracitado e o aqui apresentado estão no horizonte de previsão e na base de dados. Nesta referência, os autores, apesar de se basearem na localização das estações do INMET, utilizam dados provenientes de satélite. Em (68) os autores avaliaram a influência de fatores climáticos na geração de energia de um sistema fotovoltaico residencial on-grid em Lavras, Minas Gerais, utilizando dados de estações climatológicas convencionais para desenvolver modelos de regressão linear múltipla.

2.5 TRABALHOS DE MODELAGEM DA IRRADIAÇÃO SOLAR EM OUTROS ESTADOS DO BRASIL

Diversos trabalhos foram encontrados com tentativas de modelar a irradiação solar em regiões específicas do Brasil. Contudo, a grande maioria utiliza modelos não computacionais para tal (77, 27, 47, 90, 115, 16, 100). Também foram encontrados, em menor número, trabalhos que utilizam diferentes modelos computacionais para prever irradiação solar em diferentes locais no Brasil. Estes são analisados abaixo:

Em Soares et al. (99) um modelo **MLP** foi utilizado em uma base da cidade de

São Paulo. Para os padrões atuais é um trabalho de simples aplicação de um modelo de aprendizado de máquina, contudo à época este tipo de aplicação ainda não era tão abrangente.

Lima et al. (73) utiliza os valores de irradiação solar obtidos através de modelos de previsão numérica do tempo como entrada para uma ANN, juntamente com outras variáveis, como forma de melhorar o desempenho dos modelos de previsão numérica. A região alvo é o Nordeste, dividida em 4 áreas de acordo com a similaridade dos dados, onde um total de 110 estações foram utilizadas. A análise espacial é realizada através de modelos *kriging* para a geração de curvas de contorno dos valores previstos.

Um grupo tem ao longo dos anos trabalhado de várias formas com dados da cidade de Botucatu, estado de São Paulo. dos Santos et al. (45) investiga a eficiência de modelos computacionais, especificamente SVM e ANN, comparando-os à modelos empíricos. Os dados analisados são de Fevereiro de 1996 a Dezembro de 2008. Os modelos são utilizados com diferentes conjuntos de variáveis, selecionados aleatoriamente, objetivando previsões horárias e diárias. Algumas combinações de parâmetros apresentam resultados interessantes para ambas granularidades temporais, com o melhor resultado na previsão diária sendo $R^2 = 0.91$. da Silva et al. (35), apresenta um estudo comparativo entre o modelo empírico Angstrom-Prescott, SVM e ANN. Neste trabalho uma base de dados de 1996 a 2011 foi utilizada. O trabalho divide a base de teste considerando os anos como típicos e atípicos, segundo a média dos valores anuais de irradiação solar. Esta incomum forma de teste dos modelos apresenta resultados interessantes como R^2 de 0.887 com modelo empírico e até de 0.941 com o SVM, mas o trabalho não deixa claro se existe uma forma de caracterizar um ano em andamento como típico ou atípico. Teramoto et al. (103) estuda a previsão de irradiação solar ultravioleta em Botucatu. É feita uma análise utilizando modelos empíricos, ANN e SVM, sendo o último o que teve melhor desempenho.

Teramoto et al. (102) Compara modelos empíricos com SVM na cidade de Iguape, São Paulo. Os autores supõem que a escolha do conjunto de variáveis é o principal fator na definição da eficácia do modelo e que diferentes climas devem utilizar diferentes conjuntos de variáveis. Para tal, 8 conjuntos de variáveis são arbitrariamente definidos e utilizados em 6 modelos empíricos e SVM. Conforme esperado, o SVM apresentou melhor desempenho na estimativa da irradiação solar global diária comparado aos modelos empíricos.

dos Santos et al. (44) Compara o desempenho modelos empíricos, de regressão linear, ANN e SVM, com dados de Campo Grande, Mato Grosso do Sul, entre 2005 e 2017. Alguns subgrupos de variáveis são arbitrariamente definidos e comparados. Estranhamente os resultados obtidos mostram o melhor desempenho para modelos empíricos, seguido por SVM, modelos de regressão e ANN.

Acreditamos que esse resultado tenha vindo pela forma de calibração dos diferentes

modelos. O trabalho cita que os modelos empíricos e de regressão foram ajustados usando um software estatístico, contudo a metodologia deste software não fica clara. O modelo [SVM](#) foi treinado e validado através do Waikato Environment for Knowledge Analysis (WEKA), um software híbrido que faz o treinamento e otimização dos hiperparâmetros, enquanto o [ANN](#) aparentemente não teve os parâmetros otimizados.

de Freitas Viscondi and Alves-Souza ([39](#)) estuda o desempenho dos modelos [SVM](#), [ANN](#) e [ELM](#) na região metropolitana de São Paulo com dados entre 1962 e 2014. Alguns grupos de variáveis são definidos observando a correlação entre as variáveis contudo, na conclusão os autores observam que a utilização de todas as variáveis disponíveis foi mais eficiente. Os 3 modelos tiveram desempenhos semelhantes, com o [SVM](#) ligeiramente melhor que os demais.

Lima et al. ([74](#)) apresenta um conceito da economia chamado Teoria do Portfólio como método de *ensemble* aplicado à modelos de aprendizado de máquina para a previsão de irradiação solar na Espanha e no Brasil. Os dados do Brasil são de Fortaleza, no Ceará, entre 2007 e 2016. Semelhantemente, em Carneiro et al. ([23](#)) os autores aplicam uma técnica de *ensemble* com 4 tipos diferentes de [ANNs](#) para a predição de irradiação solar e velocidade do vento no Brasil e na Espanha. O artigo mostra uma abrangente revisão bibliográfica, para justificar a escolha dos seus métodos. Na base de dados do Brasil, as medições de irradiação solar aconteceram entre 2013 e 2016 na cidade de Petrolina, Pernambuco. Os dados vieram da base do Sistema de Organização Nacional de Dados Ambientais. Após cada modelo produzir independentemente seu resultado, é realizada uma combinação linear com os outros métodos de aprendizado para a composição do resultado final utilizando [Ridge Regression \(RR\)](#). No geral, a combinação de modelos obteve resultados melhores do que os melhores modelos aplicados individualmente.

Ferreira et al. ([49](#)) tem como objetivo a predição da evapotranspiração em todo o território do Brasil utilizando a metodologia FAO56. Para tal, se faz necessário, entre outras predições, a de irradiação solar. Quatro diferentes modelos de aprendizado de máquina foram utilizados assim como modelos estatísticos. A base de dados contém 437 estações, provenientes do [INMET](#), com medições entre 2004 e 2018. O trabalho divide o Brasil em 5 regiões climáticas homogêneas. Os modelos computacionais apresentam melhores resultados, mas não diferem muito entre si. A divisão do território em regiões também melhora os resultados encontrados comparando com a análise de todo o Brasil.

de O. Santos et al. ([40](#)) apresenta um método para seleção dinâmica do melhor modelo para previsão de irradiação solar horária em 4 cidades do Brasil, Florianópolis, Fortaleza, Salvador e São Paulo. O foco do trabalho é o método apresentado e não os dados, que são utilizados apenas como *benchmark*.

2.6 OUTROS TRABALHOS RELACIONADOS

Carvalho et al. (26) faz uma interessante análise sobre o impacto e viabilidade da utilização de energia fotovoltaica em residências de Minas Gerais. Infelizmente, os resultados ainda mostram que o custo deste tipo de tecnologia é um fator restritivo para a maioria da população. Com isso, considerando que a maior parte do público que adota este tipo de tecnologia faz parte de uma pequena parte da população, podemos imaginar quanto espaço ainda existe para a implantação de geração fotovoltaica distribuída, que provavelmente acontecerá conforme os valores forem se tornando mais acessíveis.

Antonanzas et al. (4) foca nos benefícios financeiros que uma melhor previsão da irradiação solar na geração fotovoltaica pode gerar, especificamente para as geradoras e distribuidoras de energia. O trabalho é focado no mercado de Portugal e Espanha. O estudo compara previsões por dois modelos de aprendizado diferentes e aplica os resultados à um modelo econômico para discutir os impactos da acurácia da previsão.

Visser et al. (105) compara 12 modelos que preveem a produção de energia fotovoltaica do dia seguinte considerando os requisitos do mercado holandês. Os autores analisam o efeito de agregar vários geradores distribuídos no desempenho do modelo de previsão.

Roldán-Fernández et al. (93) estima o efeito da produção residencial distribuída de energia elétrica no mercado ibérico. Os autores mostram como o crescimento da produção distribuída visando autoconsumo levará a uma redução da demanda, o que modificaria a curva de demanda, levando a uma redução tanto da energia negociada quanto do preço de compensação. Os autores concluem que o aumento em 1% da produção para autoconsumo leva a uma redução de quase 2% no custo da energia comercializada.

3 MATERIAIS E MÉTODOS

3.1 DADOS

O Instituto Nacional de Meteorologia (INMET)(65) mantém e disponibiliza dados de mais de 500 estações meteorológicas em todo o Brasil, sendo 68 delas estações meteorológicas automáticas em Minas Gerais - MG. Este trabalho utilizou dados de diferentes períodos, de dezembro de 2002 a dezembro de 2021, provenientes de 67 estações distintas em 66 municípios de todas as regiões de Minas Gerais, compondo seu banco de dados principal. A disponibilidade dos dados de cada estação depende de diversos fatores externos, como data de instalação, falhas e períodos de manutenção. Também utilizamos dados de estações em outros estados além de Minas Gerais para desenvolver a metodologia proposta neste trabalho. Para selecionar essas estações, calculamos, para cada uma das 66 cidades de Minas Gerais, as 10 estações mais próximas. Todas as cidades dessa lista foram incluídas no banco de dados deste trabalho, independentemente de estarem ou não em Minas Gerais. Foram utilizados dados de 67 estações fora de Minas Gerais: 17 do estado de São Paulo (SP), 13 do Rio de Janeiro (RJ), 12 do Espírito Santo (ES), 11 da Bahia (BA), 9 do Goiás (GO), 4 do Distrito Federal (DF) e 1 do Mato Grosso do Sul (MS).

A Tabela Suplementar S1 detalha o código de cada estação, sua cidade, unidade federativa, coordenadas geográficas, período de observação e o total de dados analisados, enquanto a Figura 4 destaca no mapa todas as cidades onde há estações meteorológicas analisadas, com cores que variam entre os estados. Neste trabalho, foram utilizadas 20 variáveis. A Tabela 1 as descreve de acordo com (65).

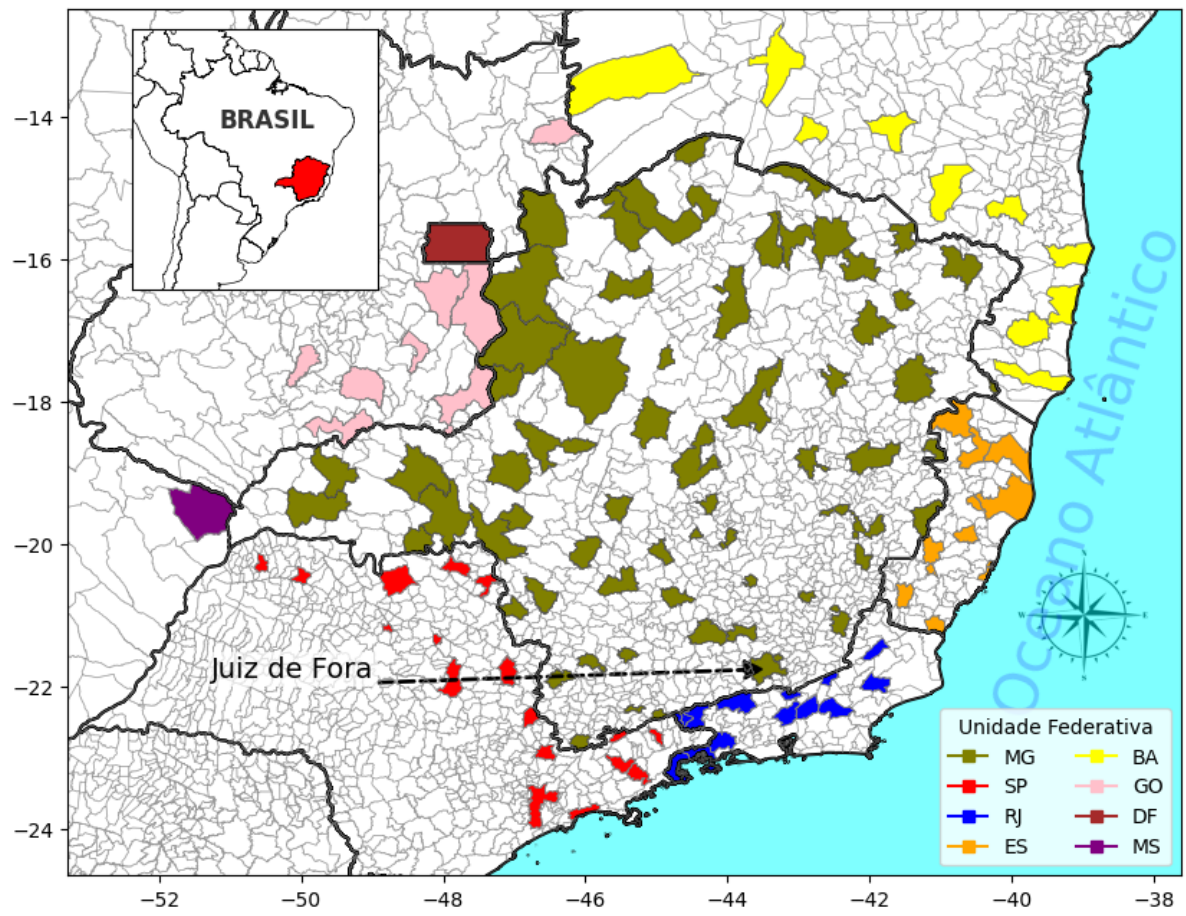


Figura 4 – Mapa das cidades estudadas

Tabela 1 – Variáveis utilizadas

Variável	Descrição
T_{med}	A temperatura média de um dia, considerando-se a média das leituras de hora em hora ou, mais frequentemente, as temperaturas máxima e mínima
T_{max}	temperatura máxima aferida em um dia
T_{min}	A temperatura mínima aferida em um dia
Ru_{med}	É a medida da umidade relativa do ar média de um dia
Ru_{max}	É o valor máximo da umidade relativa do ar de um dia
Ru_{min}	É o valor mínimo da umidade relativa do ar de um dia
Dew_{med}	A temperatura média de um dia na qual o ar teve de ser esfriado a uma pressão constante para ser saturado
Dew_{max}	A temperatura máxima em um dia na qual o ar teve de ser esfriado a uma pressão constante para ser saturado
Dew_{min}	A temperatura mínima de um dia na qual o ar teve de ser esfriado a uma pressão constante para ser saturado
P_{med}	É a medida média da pressão atmosférica ao nível da estação em um dia
P_{max}	É a medida máxima da pressão atmosférica ao nível da estação em um dia
P_{min}	É a medida mínima da pressão atmosférica ao nível da estação em um dia
$Wind$	É a medida da velocidade média do vento em um dia
$Gust$	É a média da velocidade máxima de rajada de vento em um dia
$Global$	É a medida de toda irradiação solar que chegou a superfície terrestre. Esta é a variável alvo deste trabalho. Aqui, a unidade utilizada é $MJm^{-2}d^{-1}$
Qo	É a irradiância solar extraterrestre diária. Essa é a energia disponível em um dia em dada latitude, sem se considerar os efeitos atenuantes da atmosfera
$Latitude$	É a localização de um ponto específico na superfície da Terra em relação ao equador. É medida em graus, sendo a linha do Equador zero grau. Sua representação é feita por meio de linhas paralelas que circundam o planeta horizontalmente, dividindo-o em Norte e Sul
$Longitude$	É a localização, em relação ao Meridiano de Greenwich, de um determinado ponto na superfície da Terra. Assim como a latitude, é medida em graus e o Meridiano de Greenwich corresponde a zero grau de longitude. Sua representação é feita por linhas verticais que cruzam a Terra do Polo Norte ao Polo Sul
$Altitude$	É a distância vertical entre um ponto ou objeto específico na superfície do planeta Terra e o nível médio do mar, que é uma referência fixa
$Biome$	Trata-se de um conjunto de vida vegetal e animal, constituído pelo agrupamento de tipos de vegetação próximos e identificáveis em nível regional, com condições geológicas e climáticas semelhantes e que, historicamente, passaram pelos mesmos processos de formação da paisagem, resultando em uma diversidade de flora e fauna

Vale ressaltar que os dados do INMET possuem granularidade horária. Neste trabalho, utilizamos os valores diários para cada variável, considerando que para as variáveis *Global* e *Qo*, foi utilizada a soma de cada dia, e para as demais, a média das leituras diárias. A variável *Qo*, diferentemente das outras, não é medida por dispositivos e não está disponível no INMET. Essa variável foi gerada sinteticamente pela biblioteca PySolar (1). A fonte dos valores utilizados para a variável Bioma foi fornecida pelo Instituto Brasileiro de Geografia e Estatística (64). De forma ilustrativa, a Figura 5 apresenta a

plotagem dos dados temporais para a cidade de Juiz de Fora. Nela podemos notar a relação entre a variável *Global* e a *Q0*, visto que ambas mensuram o mesmo evento físico, diferindo quanto ao local de medição, a primeira na superfície terrestre e a segunda acima da atmosfera. Também vale destacar que, assim como em outras estações, a de Juiz de Fora apresenta lacunas nas medições. No contexto deste trabalho, não realizamos nenhum tipo de imputação de dados e, sempre que alguma medição foi suspeita, ela foi removida da base.

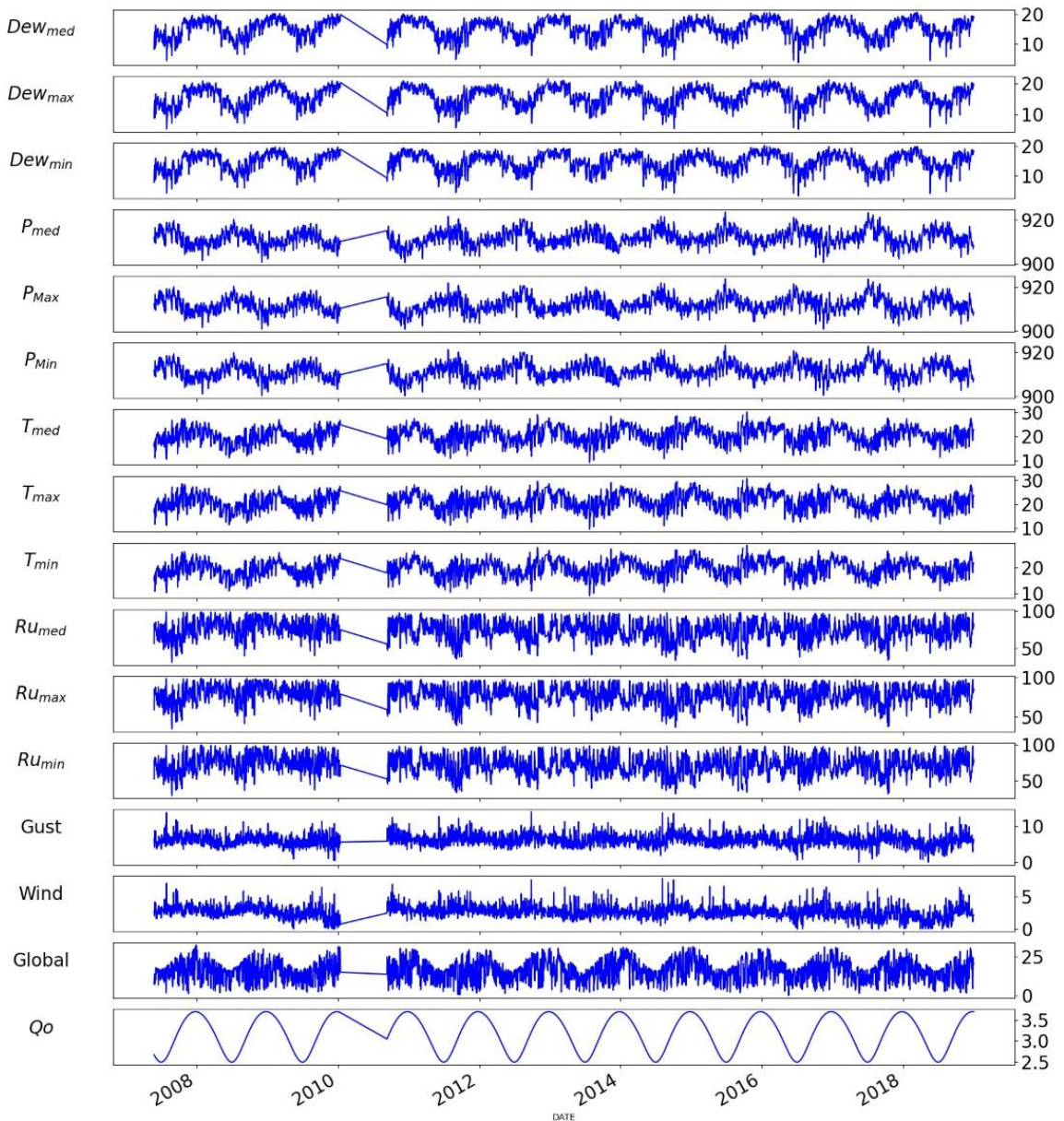


Figura 5 – Plot das variáveis temporais da cidade de Juiz de Fora

3.1.1 BIOMAS

Um bioma é uma comunidade biológica de plantas, animais e outros organismos que interagem entre si e com o ambiente físico em uma determinada região. Trata-se de uma grande área geográfica com características distintas, como clima, solo, vegetação e fauna, que a diferenciam de outras áreas circundantes.

Cada bioma abriga uma variedade única de espécies adaptadas às condições específicas do ambiente em que vivem. Essas espécies desenvolveram características particulares para sobreviver e se reproduzir nas condições daquele bioma específico. Portanto, os biomas são considerados unidades ecológicas necessárias para o estudo e a conservação da biodiversidade.

Segundo (64), o Brasil possui 6 biomas diferentes:

Amazônia: Localizada na região norte do país, ocupa 49% do território brasileiro. Este bioma está situado na maior floresta tropical do mundo, conhecida por sua biodiversidade e pela presença do Rio Amazonas. Possui vegetação densa e úmida, com diversas espécies de árvores, plantas, animais e insetos. Contém 20% da água disponível no planeta e grandes reservas minerais.

Cerrado: O Cerrado é o segundo maior bioma brasileiro, ocupando grande parte do território central. Sua vegetação é composta por árvores baixas, arbustos e gramíneas, adaptadas a solos de baixa fertilidade. Abriga uma rica diversidade de fauna e flora e é reconhecido como o Cerrado mais rico do mundo em termos de biodiversidade.

Mata Atlântica: Encontra-se principalmente na região costeira do Brasil, estendendo-se ao longo do litoral atlântico, e abriga mais de 50% da população brasileira. É um dos biomas mais ameaçados do país devido à intensa urbanização e ao desmatamento. Apenas 27% de sua cobertura florestal original ainda se conserva. A Mata Atlântica possui grande diversidade de espécies vegetais e animais e uma variedade de ecossistemas, incluindo florestas densas, manguezais e bancos de areia.

Pantanal: Localizada na região centro-oeste, é a maior área úmida do mundo e ocupa aproximadamente 2% do território nacional. Apresenta uma paisagem de planícies aluviais, rios, lagos e savanas.

Caatinga: Está presente principalmente na região Nordeste do Brasil e ocupa uma área de aproximadamente 10% do Território Nacional. É um bioma semiárido, caracterizado por vegetação adaptada à escassez de água, com árvores espinhosas, arbustos e cactos.

Pampa: Localizada na região sul do Brasil, estendendo-se até o Uruguai e a Argentina, caracteriza-se por vastas áreas de campos abertos com gramíneas e arbustos. Seu clima é chuvoso, sem período seco, mas com temperaturas negativas no inverno, que influenciam a vegetação.

Em Minas Gerais, encontramos três biomas: Cerrado, Mata Atlântica e Caatinga. Como algumas cidades estão localizadas em regiões de fronteira entre dois ou mais biomas, definimos o valor para o banco de dados como o bioma com a área predominante naquela cidade, considerando a área total da cidade e a área de cada bioma na cidade. A Figura 6 apresenta o mapa da região estudada com suas divisões geopolíticas, destacando os biomas existentes em cores diferentes.

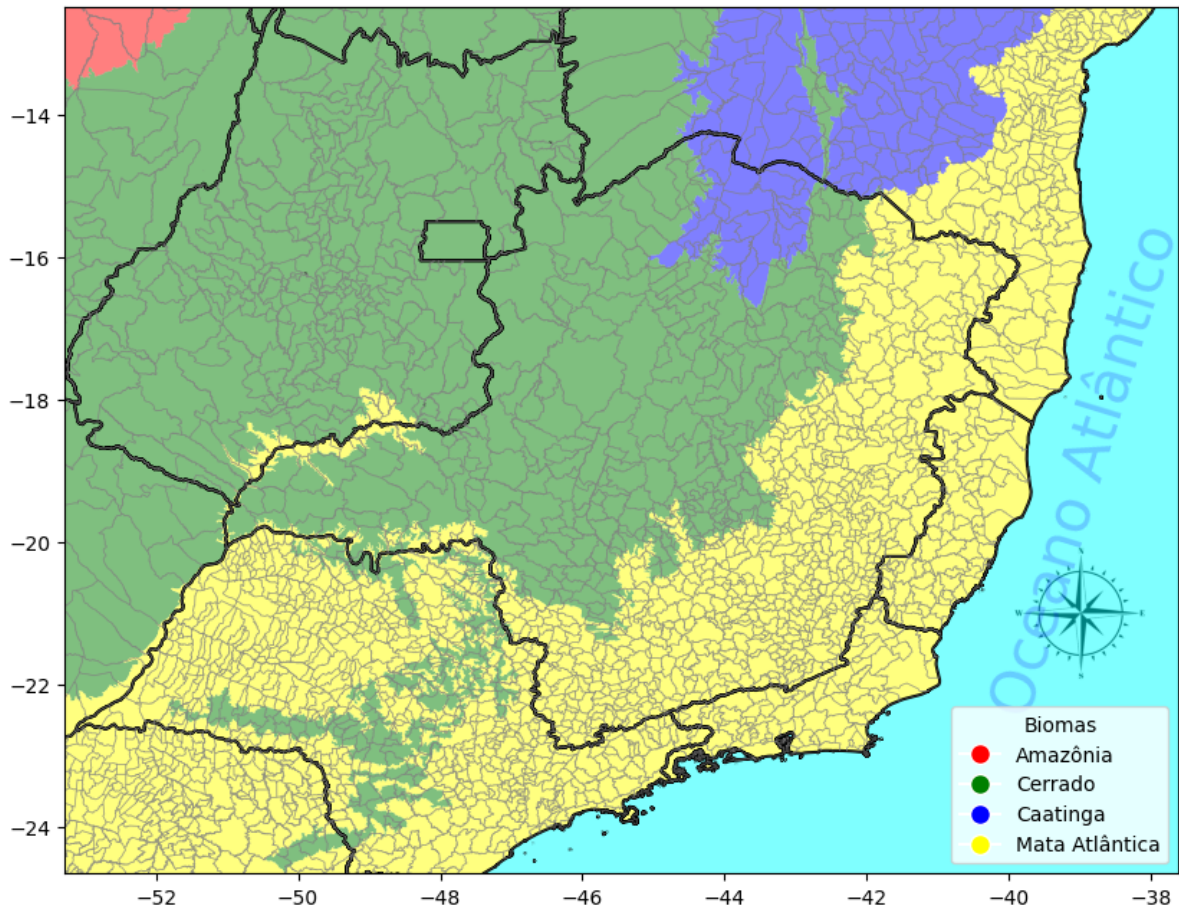


Figura 6 – Biomas dos locais estudados

3.2 FERRAMENTAL TEÓRICO

Abaixo apresentamos todos os recursos utilizados no desenvolvimento deste projeto: As métricas para a comparação dos resultados e a formulação matemática do perfis de desempenho. Também descrevemos sucintamente os modelos de aprendizado de máquina e o algoritmo de otimização aplicado.

3.2.1 MÉTRICAS DE DESEMPENHO

As métricas descritas abaixo, usadas de acordo com a documentação da biblioteca scikit-learn (87), foram usadas para avaliar o desempenho dos métodos.

Determination Coefficient:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

Mean Absolute Error:

$$MAE = \frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}_i\|$$

Mean Absolute Percentage Error:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left\| \frac{y_i - \hat{y}_i}{\max(\epsilon, \|y_i\|)} \right\|$$

Mean Bias Error:

$$MBE = \frac{1}{n} \sum_{i=1}^N (y_i - \hat{y}_i)$$

Willmott Index of Agreement:

$$WI = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (\|y_i - \bar{y}\| + \|\hat{y}_i - \bar{y}\|)^2}$$

Onde \hat{y}_i é a saída prevista, y_i é a saída medida, \bar{y} é a média dos valores medidos e N é o número total de amostras no conjunto de dados.

A Tabela 2 descreve a finalidade do uso de cada métrica, os melhores e piores resultados que podem ser encontrados e o que eles significam.

Tabela 2 – Métricas e indicadores de desempenho

Métrica	Objetivo	Melhor Resultado	Pior Resultado
R^2	Mede a proporção da variabilidade na variável dependente explicada pelas variáveis independentes	1 (perfeita)	0 (não ajustado)
RMSE	Mede o desvio médio entre os valores observados e os valores previstos, considerando todos os erros maiores com um peso maior	0 (nenhum erro)	Valores mais altos possíveis
MAE	Mede o desvio médio entre os valores observados e os valores previstos, considerando todos os erros com o mesmo peso	0 (nenhum erro)	Valores mais altos possíveis
MAPE	Mede o erro percentual médio entre os valores observados e os valores previstos.	0% (sem erro percentual)	Valores mais altos possíveis (%)
MBE	Mede os erros médios entre os valores observados e os valores previstos (valores positivos indicam superestimação, valores negativos indicam subestimação)	0 (sem viés)	Qualquer valor diferente de zero indica viés sistemático
WI	Mede a similaridade entre os valores observados e previstos, considerando tanto o erro quanto a variabilidade natural dos dados	1 (perfeita)	0 (nenhuma similaridade)

3.2.2 PERFIS DE DESEMPENHO

Os perfis de desempenho (*Performance profiles*) são um recurso valioso na avaliação comparativa de otimização. Eles fornecem um método abrangente para avaliar e contrastar a eficácia de diversos algoritmos de otimização em um espectro de cenários de teste. Quando consideramos cada estação como um problema independente, lidamos com 67 problemas individuais. Esses problemas são analisados em 5 cenários distintos, cada um sendo avaliado em 6 métricas diferentes. Conseqüentemente, essa abordagem gera um conjunto de dados abrangente de 210 resultados distintos. Gerenciar um volume tão extenso de resultados pode ser desafiador, principalmente ao lidar com métricas potencialmente inconsistentes. Para resolver isso, empregamos a técnica de avaliação de

modelos por meio de perfis de desempenho, conforme detalhado por (43). Essa abordagem facilita uma avaliação gráfica da superioridade de um solucionador sobre outro, e sua metodologia é elaborada abaixo.

Considere um conjunto P de problemas de teste p_j , com $j = 1, 2, \dots, n_p$, um conjunto A de algoritmos a_i , com $i = 1, 2, \dots, n_a$ e $t_{p,a} > 0$, e uma métrica de desempenho (como tempo de computação, média, etc.). A razão de desempenho é definida como:

$$r_{p,a} = \frac{t_{p,a}}{\min\{t_{p,a} : a \in A\}} \quad (3.1)$$

O perfil de desempenho do algoritmo é definido como:

$$\rho_a(\tau) = \frac{1}{n_p} \|\{p \in P : r_{p,a} \leq \tau\}\| \quad (3.2)$$

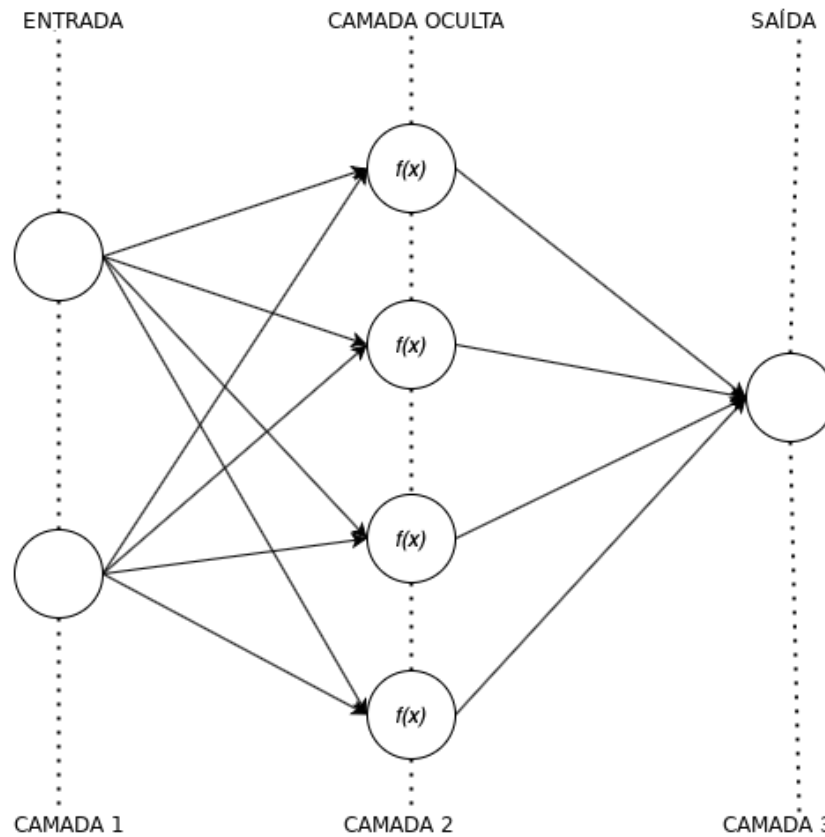
onde $\rho_a(\tau)$ é a fração de problemas resolvidos pelo algoritmo com desempenho dentro de um fator τ do melhor desempenho obtido, considerando todos os algoritmos.

3.2.3 MODELOS DE APRENDIZADO DE MÁQUINA

3.2.3.1 Artificial Neural Network

Artificial Neural Network (ANN) é um algoritmo de *deep learning* que surgiu e evoluiu a partir da ideia das redes neurais biológicas do cérebro humano. A principal abstração deste modelo é chamada de neurônio. Um neurônio recebe uma soma de valores de entrada multiplicados por seus pesos e adicionado aos seus vieses. A função de ativação é responsável por decidir se o resultado de um neurônio específico irá ou não ser utilizado nos próximos passos. Conjuntos de neurônios são agrupados em camadas e os resultados de cada camada são passados adiante como entrada para a próxima camada. A Figura 7 ilustra esta ideia.

Figura 7 – Ilustração de uma ANN



A precisão da saída de uma rede neural depende de encontrarmos os valores ideais para pesos e vieses. Tradicionalmente, os pesos e vieses são inicializados aleatoriamente e atualizados continuamente durante o processo de treinamento da rede. Considerando a equação, $y = wx$ onde w é o parâmetro de peso e x é uma entrada. Em termos simples, w define o peso dado à aquela entrada específica, x . Considerando que a solução da equação $y = wx$ sempre passará pela origem, um viés b é adicionado para melhorar o ajuste e, com isso, a equação se torna $\hat{y} = wx + b$.

3.2.3.2 Extreme Learning Machine

Extreme Learning Machine (ELM) é uma estratégia de treinamento para redes tipo Single Hidden Layer Feedforward Neural Network (SLFN) (62, 106). Este algoritmo atribui valores aleatórios aos pesos da camada de entrada, da camada oculta e dos *bias* na camada oculta. Estes parâmetros são congelados durante o processo de treino. As funções de ativação não linear na camada oculta fornecem não linearidade para o sistema. Com isto, o processo pode ser considerado como um sistema linear. O único parâmetro que precisa ser descoberto é o peso entre a camada oculta e a camada de saída. Assim, ELM converge muito mais rápido do que algoritmos tradicionais porque aprende sem iteração enquanto nós ocultos aleatórios proporcionam uma capacidade de aproximação

generalizada.

Formalmente, considerando um conjunto de testes $S = \{(x_i, t_i) | x_i = (x_{i1}, x_{i2}, \dots, x_{im})^T \in R^n, t_i = (t_{i1}, t_{i2}, \dots, t_{im})^T \in R^m\}$, onde x_i é um valor de entrada e t_i um valor alvo, a saída o de uma rede **ELM** com \hat{N} neurônios escondidos pode ser expressa como:

$$\sum_{i=1}^{\hat{N}} \beta_i g(w_i, x_j + b_i) = o_j, j = 1, \dots, N \quad (3.3)$$

Onde $g(x)$ é a função de ativação da camada oculta, N é o número de amostras de treinamento, i e j são os índices das amostras de treinamento e dos nós da camada de saída. Pode-se provar que o **SLFN** é capaz de prever perfeitamente todas as amostras de treino quando o número de nós ocultos N se aproxima do infinito. Logo, deve haver um conjunto w_i , b_i e β_i onde:

$$\sum_{i=1}^{\hat{N}} \beta_i g(w_i, x_j + b_i) = t_j, j = 1, \dots, m \quad (3.4)$$

Desta forma, treinar uma rede **SLFN** é encontrar os melhores valores de w_i , b_i e β_i .

3.2.3.3 Elastic Net

Elastic Net (EN) é um modelo de regressão linear (52) que calcula os coeficientes de um modelo linear $w = [w_0, w_1, w_2, \dots, w_p]$ enquanto que minimiza a soma dos quadrados dos resíduos entre valores preditos e mensurados. A combinação entre a regularização normativa dos coeficientes L_1 e L_2 permitem o aprendizado em um modelo esparsos onde poucos dos pesos são diferentes de zero como o modelo Lasso, mantendo as propriedades de regularização do modelo Ridge. Sendo assim, o modelo linear generalizado pode ser descrito como:

$$\min_w \frac{1}{2n_{samples}} \|Xw - y\|_2^2 + \alpha \rho \|w\|_1 + \frac{\alpha(1 - \rho)}{2} \|w\|_2^2 \quad (3.5)$$

Onde X é a matriz de pontos de dados, y são os vetores de saída, $\alpha \leq 0$, $\|w\|_1$ e $\|w\|_2$ são respectivamente a norma L_1 e L_2 do vetor de parâmetros e ρ é a razão do parâmetro L_1 .

3.2.3.4 Multivariate Adaptive Regression Spline

Multivariate Adaptive Regression Spline (MARS) é um método de regressão flexível que procura automaticamente por interações e relacionamentos não lineares (53, 30). O modelo constrói esta relação a partir de um conjunto de coeficientes e funções de base direcionadas a partir dos dados de regressão. De certa forma, o método é baseado na

estratégia “dividir e conquistar”, que divide o espaço de entrada em regiões, cada uma com sua própria equação de regressão. Sendo assim, o modelo é definido por:

$$\hat{y}(x) = F_m(x) = c_0 + \sum_{m=1}^M c_m B_m^K(x) \quad (3.6)$$

Onde c_0 é um valor constante, c_m é o coeficiente associado à m -ésima função de base, M é o número de funções de base, e K é o número de divisões, q é a ordem do polinômio, $B_m^K(x)$ é a m -ésima função de base que tem a forma:

$$B_m^K(x) = \prod_{k=1}^K [\pm(x - t)]_+^q \quad (3.7)$$

3.2.3.5 Extreme Gradient Boosting

Extreme Gradient Boosting (XGB) é uma implementação de *gradient boosting* que tem bom desempenho sendo utilizada para aprendizado supervisionado (63, 29). O XGB funciona da seguinte forma: Considerando um conjunto de dados que possui m características e um número n de amostras $(x_1, y_1), \dots, (x_n, y_n)$ onde $x_i \in \mathbb{R}^n$ e $y_i \in \mathbb{R}$, $i = 0, \dots, n$. Seja \hat{y}_i a saída estimada de um método *ensemble tree* obtido a partir das equações:

$$\phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathbb{F} \quad (3.8)$$

onde K representa o número de árvores no modelo, f_k representa a k -ésima árvore de decisão. A árvore de decisão f_k tem sua profundidade menor ou igual a m_{depth} . Para construir o modelo é necessário encontrar o conjunto de funções minimizando a perda e o objetivo de regularização.

$$L(\phi) = \sum_i l(y_i, \phi(x_i)) + \frac{1}{2} \|w\|^2 \quad (3.9)$$

onde $l = |\hat{y}_i - y_i|$ é a função de perda, sendo \hat{y}_i a saída prevista e y_i a saída real, T é o número de folhas da árvore e w é o peso de cada folha.

3.2.3.6 Support Vector Regression

Support Vector Regression (SVR) (98) é um modelo de regressão que usa funções *kernel* lineares ou não lineares para aproximar as amostras no conjunto de dados. O objetivo do **SVR** é minimizar os coeficientes, especificamente a norma L_2 do vetor de coeficientes. O termo de erro é tratado nas restrições, onde definimos o erro absoluto menor ou igual a uma margem especificada ϵ , chamada de erro máximo.

Dado um conjunto $(x_1, y_1), \dots, (x_l, y_l)$, o **SVR** resolve o seguinte problema de otimização

$$\min \frac{1}{2}(\alpha - \alpha^*)^T K(\mathbf{x}_i, \mathbf{x}_j)(\alpha - \alpha^*) + \sum_{i=1}^l (y_i + \varepsilon)(\alpha_i - \alpha_i^*) \quad (3.10)$$

sujeito a

$$\begin{aligned} e^T(\alpha - \alpha^*) &= 0, \\ 0 &\leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l. \end{aligned}$$

onde $x_i \in R^n$ é um vetor de dados de entrada, $y_i \in R^1$ é o vetor de valores observados, $\varepsilon > 0$ e $C > 0$ são os parâmetros do modelo, $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$, e $\phi(\cdot)$ é a função *kernel*.

Resolver a Eq. (3.10) permite determinar os parâmetros para construir a aproximação **SVR**. Então, as estimativas de **SVR** são dadas por

$$\hat{y}_i = \varepsilon \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b.$$

3.2.4 ALGORITMOS DE OTIMIZAÇÃO

3.2.4.1 Simple Genetic Algorithm

O **Simple Genetic Algorithm (SGA)** é um método para resolver problemas de otimização com e sem restrição inspirado seleção natural. Cada possível solução do problema é considerado um indivíduo da população e o algoritmo modifica repetidamente uma população em busca de indivíduos mais aptos. A cada passo, o algoritmo seleciona indivíduos da população atual para serem pais e os usa para produzir os filhos para a próxima geração. Ao longo de sucessivas gerações, a população evolui em direção a uma solução ótima. Os principais passos deste algoritmo estão apresentados no Algoritmo 1.

Algoritmo 1 Um algoritmo genético simples

pop ← Inicie uma população aleatória com *N* indivíduos;

repita

pop2 ← *N* melhores indivíduos de *pop*; \\ com repetição permitida

pop3 ← uma nova população com *N* indivíduos obtidos aplicando crossover a *pop2*;

pop4 ← uma nova população com *N* indivíduos obtidos aplicando a mutação em *pop3*;

 Avalie todos os novos cromossomos em *pop4*;

pop ← os melhores *N* indivíduos retirados de *pop* e *pop4*;

até *i* < *gen*

3.2.5 TRANSFERÊNCIA DE APRENDIZADO

A transferência de aprendizado (*Transfer Learning*) é uma técnica de aprendizado de máquina que envolve a utilização do conhecimento adquirido em uma tarefa ou domínio para melhorar o desempenho em uma tarefa ou domínio diferente, porém relacionado. Na transferência de aprendizado, um modelo pré-treinado, geralmente treinado em um grande conjunto de dados, é usado como ponto de partida para uma nova tarefa, em vez de treinar um modelo do zero.

A ideia por trás da transferência de aprendizado é que o conhecimento adquirido por um modelo durante o aprendizado de uma tarefa pode ser aproveitado para acelerar o aprendizado ou melhorar a generalização em uma tarefa diferente. Ao começar com um modelo pré-treinado, o modelo já possui características, padrões ou representações aprendidas que são geralmente úteis em diversas tarefas. Essas características aprendidas podem ser utilizadas como base para a nova tarefa, permitindo que o modelo se adapte e se especialize mais rapidamente.

O processo de transferência de aprendizado normalmente envolve as seguintes etapas:

Pré-treinamento: Um modelo é treinado em um grande conjunto de dados proveniente de uma tarefa ou domínio de origem. Essa etapa de treinamento geralmente é computacionalmente dispendiosa e demorada.

Extração de características: O modelo pré-treinado é usado para extrair características ou representações relevantes dos dados da tarefa de origem. Essas características capturam padrões ou informações importantes nos dados.

Ajuste fino: As características extraídas são então usadas para inicializar um novo modelo projetado especificamente para a tarefa ou domínio de destino. Esse novo modelo é treinado em um conjunto de dados menor, específico para a tarefa de destino, que geralmente é rotulado ou anotado.

Adaptação: O novo modelo é treinado novamente no conjunto de dados da tarefa de destino, geralmente com uma taxa de aprendizado menor, para adaptar os parâmetros do modelo às especificidades da tarefa de destino. Essa etapa permite que o modelo ajuste suas características aprendidas e melhore seu desempenho na tarefa de destino.

A transferência de aprendizado pode ser particularmente benéfica quando a tarefa de destino possui dados limitados disponíveis, pois ajuda a mitigar o risco de sobreajuste e melhora a capacidade de generalização do modelo. Tem sido aplicado com sucesso em diversos domínios, incluindo visão computacional, processamento de linguagem natural e análise de áudio, permitindo o desenvolvimento de modelos mais precisos e eficientes.

3.2.6 SELEÇÃO DE CARACTERÍSTICAS AUTOMATIZADA

O processo de seleção de características é crucial na construção de modelos de aprendizado de máquina. Esse processo implica selecionar as variáveis ou características mais relevantes de um conjunto de dados. É essencial para reduzir a dimensionalidade, aprimorar a interpretabilidade do modelo e melhorar o desempenho geral(59).

O desenvolvedor pode realizar o processo de seleção de variáveis arbitrariamente, utilizando as variáveis consideradas atrativas e descartando as demais. Nesse processo, geralmente observa-se a matriz de correlação das variáveis. Considerando o número de diferentes bancos de dados e suas combinações, essa abordagem seria extremamente demorada e ineficiente para o presente trabalho. Dessa forma, o processo de seleção de recursos foi realizado automaticamente por um algoritmo de otimização, juntamente com a otimização dos hiperparâmetros do modelo de aprendizado de máquina. Portanto, todas as variáveis disponíveis foram inicialmente consideradas utilizáveis no treinamento do modelo final. No entanto, o processo de otimização selecionou apenas aquelas que apresentaram o melhor desempenho. Os detalhes desse processo são descritos abaixo na seção 3.3.1.

3.3 METODOLOGIA

3.3.1 ABORDAGEM COMPUTACIONAL HÍBRIDA

Visando um parâmetro de comparação eficiente para o modelo geolocalizado, criamos um modelo computacional pontual para cada estação em Minas Gerais. Ao tratar de uma estação específica individualmente, a denominaremos Estação Alvo. A metodologia aplicada para o desenvolvimento desses modelos já foi amplamente validada em (11, 12, 13) e é descrita a seguir: Inicialmente, todos os modelos de aprendizado de máquina são executados individualmente para encontrar o melhor conjunto de hiperparâmetros para cada modelo aplicado a cada estação. Cada modelo de aprendizado de máquina possui um conjunto de valores contendo os limites superior e inferior de seus hiperparâmetros. Nesse momento inicial, é possível optar por usar um subconjunto específico de variáveis, todas as variáveis disponíveis ou aplicar uma técnica de seleção de características. Após as configurações iniciais, o algoritmo de otimização gera aleatoriamente uma população de soluções candidatas. Cada solução candidata representa um conjunto de hiperparâmetros associados ao modelo de aprendizado de máquina. Se o processo de seleção de características for realizado automaticamente, cada solução terá seu próprio conjunto de variáveis. A solução é avaliada por meio de uma estratégia de validação cruzada k-fold, onde o objetivo a ser minimizado é o valor da métrica **RMSE**, calculado entre os valores observados e previstos. Como trabalhamos com séries temporais, é importante observar que, nas dobras do processo de validação cruzada, os valores futuros não são usados para treinamento. Esse processo é conhecido como Validação Cruzada por Divisão de Séries Temporais. Quando o

critério de parada é atingido, o ciclo evolutivo termina e a solução que apresenta o melhor **RMSE** é armazenada. Com os modelos já executados, podemos analisar qual teve o melhor desempenho na Estação Alvo e conhecer a melhor configuração para seus hiperparâmetros. A Figura 8 ilustra esse processo.

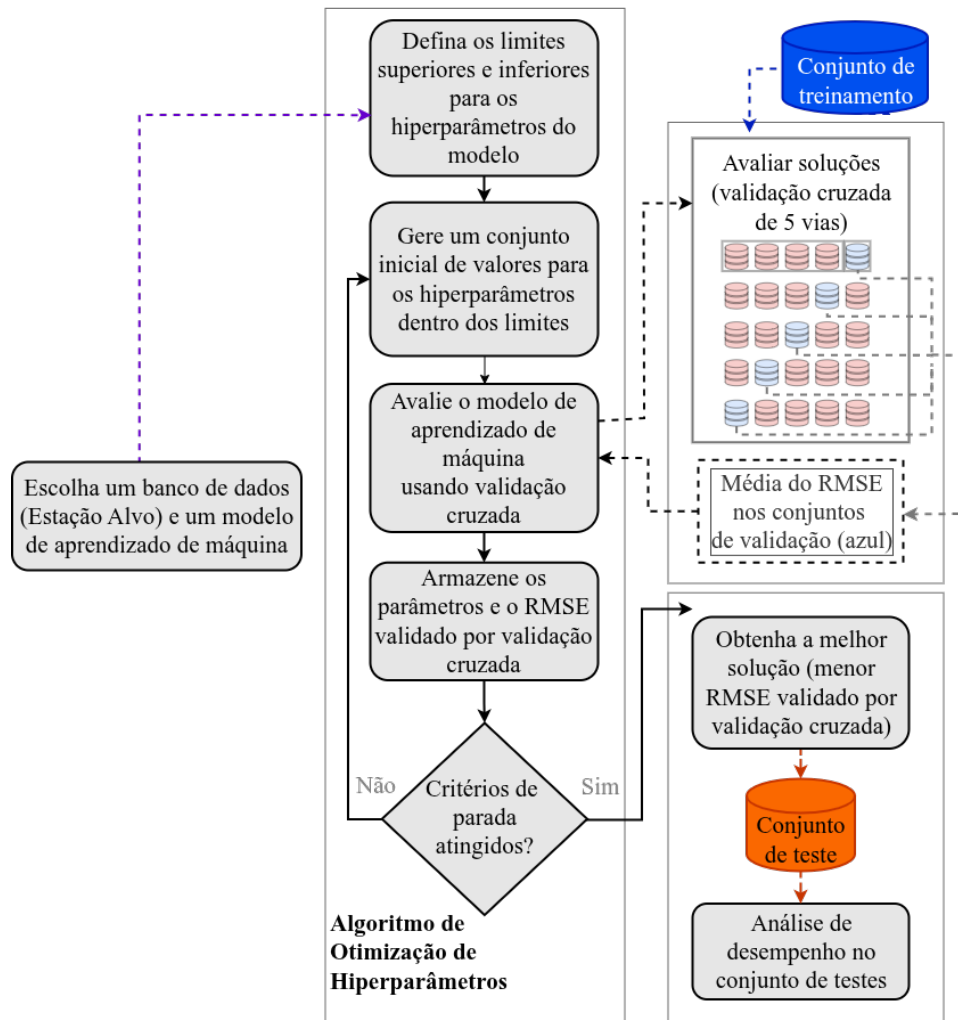


Figura 8 – Processo de otimização para cada modelo

3.3.2 MODELO GEOLOCALIZADO

Para desenvolver um modelo computacional georreferenciado de Minas Gerais, os dados das N estações mais próximas da Estação Alvo são agrupados em uma única base com os dados de geolocalização de todas as estações envolvidas.

A metodologia descrita na Seção 3.3.1 é então aplicada a este novo banco de dados. No entanto, o processo de validação ocorre apenas com os dados da Estação Alvo. Dessa forma, o modelo que representa a Estação Alvo também terá os dados das estações vizinhas em sua formulação e poderá ser considerado um modelo geolocalizado. A Figura 9 mostra essa dinâmica, enquanto a Figura 10 ilustra essa etapa no mapa. É necessário

um cuidado especial com a divisão entre os intervalos de treinamento e teste. Algumas estações, como CARATINGA e ESPINOSA, não possuem dados até o final do período normalmente observado, 31/12/2021. Portanto, os dados das estações vizinhas a essas bases também devem ser descartados na data da última leitura dessas estações quando elas forem a Estação Alvo.

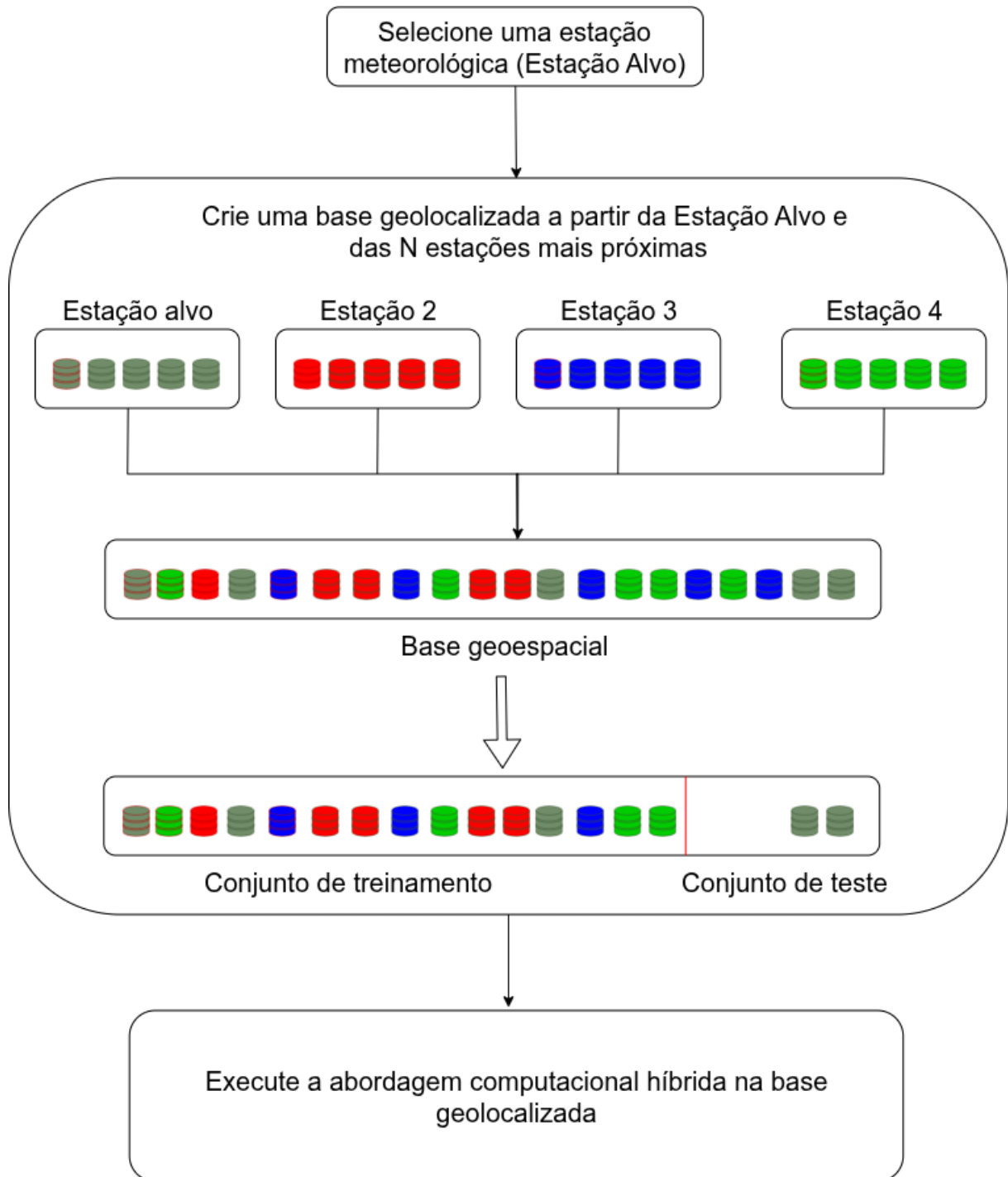


Figura 9 – Formação da base geolocalizada

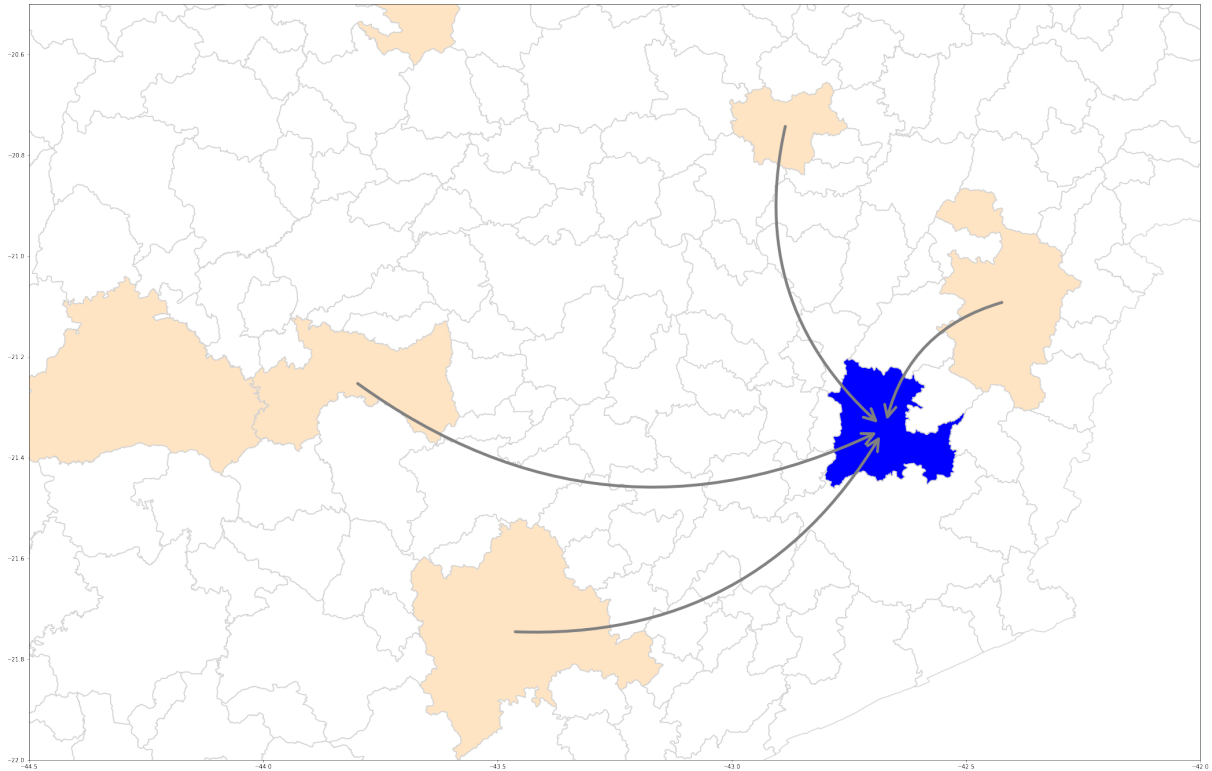


Figura 10 – Agrupamento de cidades vizinhas

Para verificar a viabilidade de prever a irradiação solar em locais sem estações meteorológicas, aplicamos uma abordagem diferente, modificando ligeiramente a metodologia descrita anteriormente. Criamos um banco de dados com dados das N estações mais próximas da Estação Alvo. No entanto, removemos os dados da Estação Alvo da base de treinamento. Dessa forma, todo o treinamento é realizado com dados de estações vizinhas, e a validação é feita com dados da Estação Alvo. Esse processo é, em essência, o que chamamos de transferência de aprendizado. A Figura 11 mostra o processo de criação da base geoespacial sem a presença da Estação Alvo na base de treinamento.

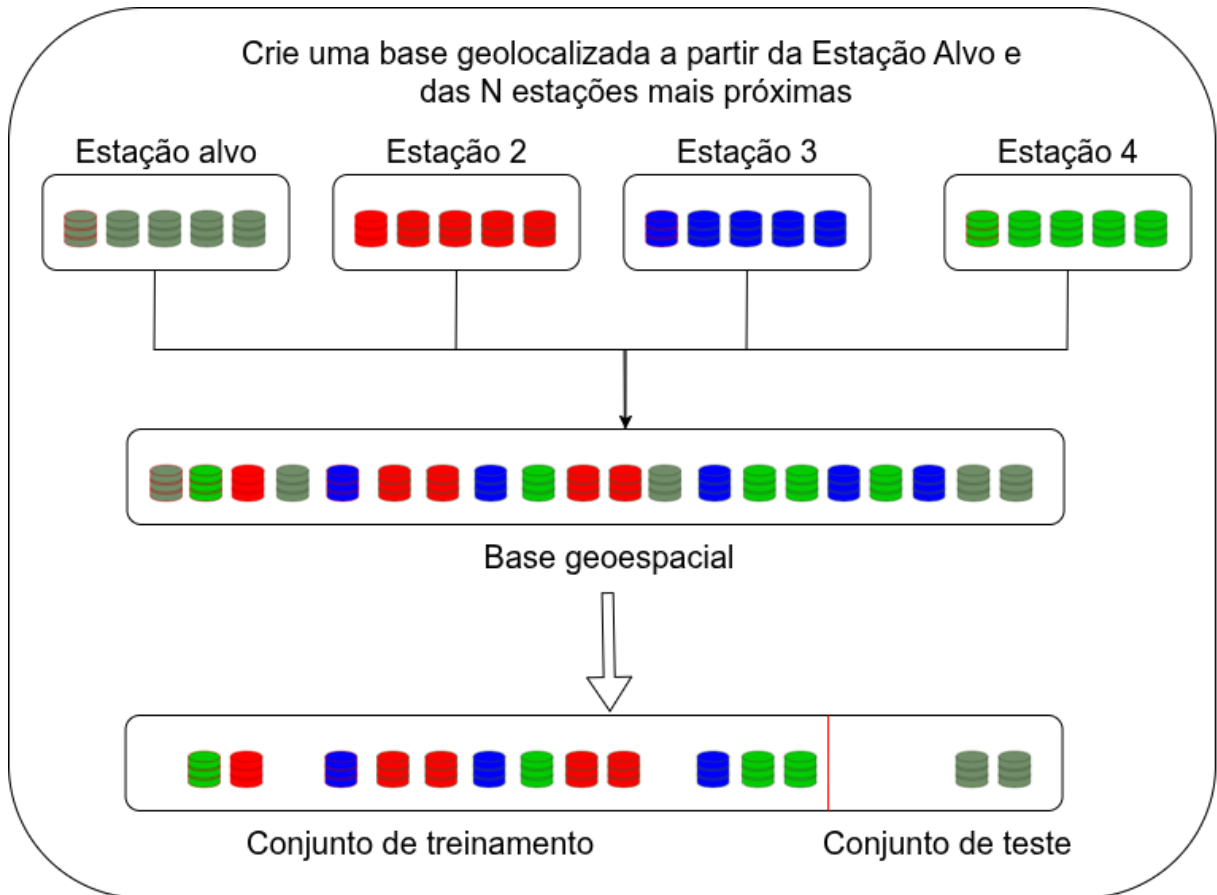


Figura 11 – Criação de banco de dados para o processo de transferência de aprendizado

3.3.2.1 SELEÇÃO AUTOMÁTICA DAS BASES DE TREINO

Também realizamos uma modificação do processo descrito acima quanto à quantidade e forma de escolha das estações que compõem a base de treino. Nesta abordagem, a quantidade de cidades que comporão a base de treino não é previamente definida. Todo o processo que define quais cidades farão parte da base de treino é realizado pelo algoritmo otimizador juntamente com o processo de otimização dos hiperparâmetros e seleção de características.

4 EXPERIMENTOS COMPUTACIONAIS, ANÁLISE E DISCUSSÃO

Embora a metodologia descrita previamente possa ser utilizada em qualquer horizonte de previsão, neste trabalho, a previsão é feita com um dia de antecedência, observando-se o dia anterior. Como os dados de cada estação têm diferentes períodos de disponibilidade em cada combinação de Cenário e Estação Alvo, diferentes períodos foram utilizados para os testes e treinamentos do modelo de aprendizado de máquina.

Realizamos três etapas de experimentos, sempre buscando reduzir a exploração e aprimorar exploração. Na primeira etapa testamos a metodologia proposta com seis diferentes modelos de aprendizado de máquina. Na segunda etapa escolhemos o modelo com o melhor desempenho e estudamos a eficácia da inserção de dados geolocalizados na base de treino. Além disso, também analisamos o comportamento do modelo na aplicação de transferência de aprendizado. Por fim, na terceira etapa, delegamos ao otimizador a tarefa de escolher o melhor conjunto de cidades para composição da base de treino para o processo de transferência de aprendizado. A Figura 12 ilustra estes processos. Na primeira e segunda etapa, cada experimento foi executado de forma independente 30 vezes. Na terceira etapa, cada experimento foi executado 10 vezes. Na apresentação dos resultados mostraremos as médias destas execuções. As Tabelas 3 e 4 apresentam, respectivamente, os parâmetros do algoritmo de otimização e os intervalos dos hiperparâmetros dos modelos de aprendizado de máquina.

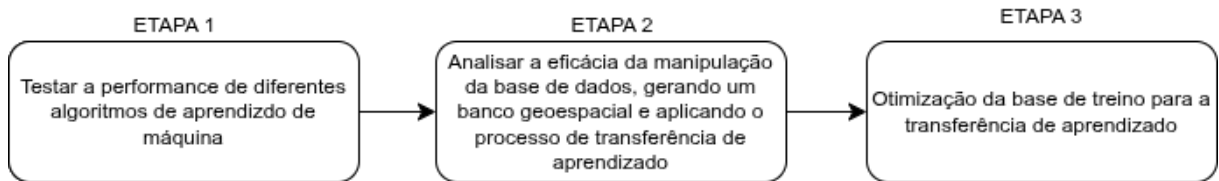


Figura 12 – Etapas de experimentos realizados

Tabela 3 – Parâmetros do algoritmo de otimização.

Algoritmo	Parametro	Descrição	Configurações
SGA	x_1	Número de gerações	[30]
	x_2	Fator de cruzamento	[0.95]
	x_3	Tamanho da população	[20]
	x_4	Probabilidade de mutação	[0.2]
	x_5	Estratégia para o cruzamento	[<i>single</i>]
	x_6	Estratégia para a mutação	[<i>uniform</i>]

Tabela 4 – Parametros utilizados para os modelos de aprendizado de máquina.

Modelo	Parametro	Descrição	Configurações/ Intervalo
MARS	x_1	Termo de penalização	[1,9]
	x_2	Número máximo de termos	[1,200]
	x_3	O grau máximo de interação	[0,2]
ANN	x_1	Número de camadas	[1, 3]
	x_2	Número de neurônios por camada	[1, 100]
	x_3	Função de ativação	[<i>Logistic</i> , <i>Identity</i> , <i>ReLU</i> , <i>Tanh</i>]
	x_4	Coefficiente de regularização L_2	[$1e^{-8}$, 1]
	x_5	Solver	[<i>adam</i>]
	x_6	Taxa de Aprendizagem	[<i>adaptive</i>]
	x_7	Taxa de Aprendizagem Inicial	[$1e^{-6}$,1.0]
	x_8	Número máximo de iterações	[1100]
ELM	x_1	Número de neurônios por camada	[1,300]
	x_2	Função de ativação	[<i>Identity</i> , <i>Gaussian</i> , <i>Multiquadric</i> , <i>Inv_Multiquadric</i> , <i>ReLU</i> , <i>Swish</i> , <i>Sigmoid</i>]
	x_3	Coefficiente de regularização α	[1.00000,1.00001]
	x_4	Largura da função de base radial	[1,10]
	x_5	Coefficiente de regularização L_2	[0,10000]
EN	x_1	Termo de penalização	[$1e^{-6}$,2]
	x_2	Coefficiente de regularização L_1	[0,1]
XGB	x_1	Taxa de Aprendizagem	[$1e^{-6}$,1]
	x_2	Número máximo de termos	[10,100]
	x_3	Profundidade máxima	[1,20]
	x_4	Coefficiente de regularização λ	[0,100]
SVR	x_1	Coefficiente de regularização γ	[0,10]
	x_2	Coefficiente de regularização C	[0.1,10000]
	x_3	Coefficiente de regularização ϵ	[0.1,100]

4.1 PRIMEIRA ETAPA DE EXPERIMENTOS

No processo de desenvolvimento do estudo aqui relatado, diferentes abordagens foram testadas. Inicialmente o framework desenvolvido não levava em consideração o processo de transferência de aprendizado. Nesta etapa, alguns experimentos foram realizados utilizando-se diferentes modelos de aprendizado de máquina. Relataremos a seguir os resultados encontrados nesta etapa com o intuito de justificar algumas escolhas feitas no processo final.

Nesta etapa preliminar, também analisamos a correlação do bioma com os resultados. Inicialmente, presumimos que essa variável apresentaria forte correlação, visto que o bioma é altamente influenciado por diversos fatores climáticos, como altitude e geografia geral da região. Contudo, na maioria dos testes realizados, o processo de seleção de variáveis excluiu essa variável exógena, demonstrando que ela não influenciou os resultados. Portanto, optamos por não utilizar esses dados na segunda fase dos experimentos.

Inicialmente, foram aplicados seis diferentes modelos de aprendizado de máquina, seguindo a metodologia descrita em (13). Os modelos foram **Artificial Neural Networks** (ANN), **Extreme Learning Machine**(62, 106) (ELM), **Elastic Net**(52) (EN), **Multivariate Adaptive Regression Spline**(53, 30) (MARS), **Extreme Gradient Boosting**(63, 29) (XGB), e **Support Vector Regression**(98) (SVR).

De todos os modelos aplicados o que demonstrou melhor desempenho foi o ELM, conforme podemos observar na Tabela 5, que mostra a área normalizada sob as curvas dos perfis de desempenho de cada modelo de aprendizado.

Tabela 5 – Áreas normalizadas dos perfis de desempenho - Etapa 1

Modelo de aprendizado de máquina	Área
ANN	0.904
ELM	1.000
EN	0.911
MARS	0.996
SVR	0.799
XGB	0.979

4.2 SEGUNDA ETAPA DE EXPERIMENTOS

Considerando que o modelo ELM apresentou os melhores resultados nos testes iniciais, nós o aplicamos exclusivamente na segunda fase de desenvolvimento. Este algoritmo de aprendizado de máquina é, especificamente, uma variação da rede neural *feedforward*, e seu treinamento é feito de forma diferente da maioria das outras redes neurais. É um dos modelos de aprendizado de máquina mais utilizados para prever a irradiação solar(6).

Ao contrário das redes neurais tradicionais, que utilizam um processo de treinamento

iterativo e gradual para ajustar os pesos das conexões entre as camadas, o ELM atribui esses pesos aleatoriamente. O ELM não requer uma fase de treinamento longa e demorada, tornando-o muito mais rápido do que outros métodos de aprendizado de máquina.

O ELM também é conhecido por ser altamente eficiente em termos de processamento e requer relativamente poucos dados de treinamento para produzir resultados precisos. É frequentemente usado em tarefas de classificação e previsão, como reconhecimento de padrões, análise de séries temporais e processamento de sinais.

Como otimizador, utilizamos um Algoritmo Genético Simples (AG). O AG é um método para resolver problemas de otimização com e sem restrições, inspirado na seleção natural. Cada solução possível para o problema é considerada um indivíduo da população, e o algoritmo modifica repetidamente a população em busca de indivíduos mais aptos. A cada passo, o algoritmo seleciona indivíduos da população atual como pais e os utiliza para gerar descendentes para a próxima geração. Ao longo das gerações sucessivas, a população evolui em direção a uma solução ótima. A implementação utilizada foi a biblioteca Pygmo(18).

Todos os cenários apresentados foram executados 30 vezes de forma independente, visando maior confiabilidade estatística nos resultados encontrados. Cinco cenários diferentes foram analisados para todas as estações:

- 1 - O banco de dados da Estação Alvo foi utilizado isoladamente para treinamento e teste. Este cenário representa o modelo da estação de forma independente, é o tradicionalmente analisado em outros trabalhos (33, 13) e serve como principal objeto de comparação dos resultados.
- 2 - Os dados das duas cidades vizinhas mais próximas foram adicionados às bases de treinamento da Estação Alvo. Apenas os dados da Estação Alvo foram usados como base de teste. Nesse cenário, verificamos o benefício de usar dados de estações vizinhas para compor a base de treinamento. Podemos considerar os modelos gerados nesse cenário como geolocalizados.
- 3 - Os dados das duas cidades vizinhas mais próximas foram usados como base de treinamento, e os dados da Estação Alvo como base de teste. Este cenário busca verificar a viabilidade da aplicação de transferência de aprendizado, utilizando cidades com estações meteorológicas como base de treinamento, para gerar modelos para cidades sem estações meteorológicas.
- 4 - Os dados das quatro cidades vizinhas mais próximas foram adicionados às bases de treinamento da Estação Alvo. Apenas os dados da Estação Alvo foram usados como base de teste. Este cenário é semelhante ao cenário 2, com a única diferença sendo o número de estações utilizadas.

5 - Os dados das quatro cidades vizinhas mais próximas foram usados como base de treinamento e os dados da Estação Alvo como base de teste. Este cenário é semelhante ao cenário 3, com a única diferença sendo o número de estações utilizadas.

A Tabela Suplementar S2 apresenta as médias de 30 simulações independentes para todas as cidades nos cinco cenários.

4.2.1 DISCUSSÃO

Considerando o cenário 1 e comparando-o com os principais trabalhos encontrados na literatura (33, 13), apenas neste trabalho são estudadas 16 estações. Em 22 estações, os resultados apresentados aqui são melhores. Nas outras 29 estações, os resultados relatados na literatura são melhores, sempre observando a métrica R^2 . Esses resultados podem indicar a superioridade do modelo MARS sobre o ELM. No entanto, os experimentos apresentados aqui não são suficientes para comprovar essa situação. Essa variação nos resultados pode ser gerada por diferentes razões, como o intervalo observado, a porcentagem de divisão entre as bases de treinamento e teste e a diferença entre os modelos empíricos e de aprendizado de máquina utilizados em outros trabalhos. De qualquer forma, este trabalho visa comprovar a viabilidade do uso de dados geograficamente distantes para prever valores sem estações meteorológicas e não obter modelos melhores onde já existem estações. A comparação dos resultados deste cenário com os encontrados na literatura pode ser vista na Tabela Suplementar S3, com os melhores resultados de cada cidade destacados em negrito.

Por meio dos cenários 2 e 4, tentamos verificar a hipótese de que a adição de dados de estações vizinhas melhoraria os resultados em um ponto específico. Constatamos que, para a maioria das estações, precisamente 47, a adição desses dados à base de treinamento não foi vantajosa. Ou seja, em 47 estações, o resultado do cenário 1 foi melhor do que os resultados dos cenários 2 e 4. Em 11 estações, o cenário 2 apresentou melhores resultados e, em 9, o cenário 4. Sempre observando a métrica R^2 .

Os cenários 3 e 5 apresentam a hipótese central deste trabalho, onde aplicamos a transferência de aprendizado para prever valores de irradiação solar em cidades sem estações meteorológicas. Para isso, como explicado anteriormente, consideramos cada cidade como se não possuísse uma estação meteorológica e utilizamos dados de cidades vizinhas para treinar o modelo. Utilizamos apenas dados da Estação Alvo para confirmar a eficiência da transferência de aprendizado no período de validação do modelo.

Especificamente no cenário 3, em 39 das 67 cidades, houve execuções independentes em que a métrica R^2 foi negativa. Em 16 delas, os resultados médios dessa métrica foram menores que zero. Nas outras 23, mesmo com algumas execuções ruins, a média final foi razoável. Nesse cenário das 2010 execuções independentes (67×30), 234 apresentaram a métrica R^2 menor que zero, em torno de 11,65%.

No cenário 5, em 11 cidades, houve execuções independentes onde a métrica R^2 foi negativa e, em 4, os resultados médios dessa métrica foram menores que zero. Analisando o total de execuções, neste cenário, das 2010 execuções independentes (67×30), 83 apresentaram a métrica R^2 menor que zero, em torno de 4,13%.

A Figura 13 mostra os diagramas de dispersão da irradiação solar medida e estimada para as melhores execuções individuais de cada estação, considerando a métrica R^2 , agrupadas por cenários. A Figura 14 mostra o gráfico dos perfis de desempenho para os cinco cenários, considerando todas as cinco métricas utilizadas. A Tabela 6 mostra os valores das curvas dos perfis de desempenho normalizados. Analisando esses resultados, apesar das curvas dos perfis de desempenho dos cinco cenários serem muito próximas e os diagramas de dispersão muito semelhantes, podemos confirmar que, para este estudo, o uso de dados de cidades vizinhas na configuração da base de treinamento não foi vantajoso, visto que a curva do cenário 1 cresceu mais rapidamente do que as curvas dos cenários 2 e 4. Em relação ao uso de transferência de aprendizado para previsão de dados em locais sem estações meteorológicas, a análise por meio dos perfis de desempenho confirma que, para este estudo, o aumento no número de cidades utilizadas gera melhor desempenho. Mais importante ainda, a proximidade das curvas dos cenários 3 e 5 com a curva do cenário 1 indica que essa técnica pode ser usada de forma eficiente.

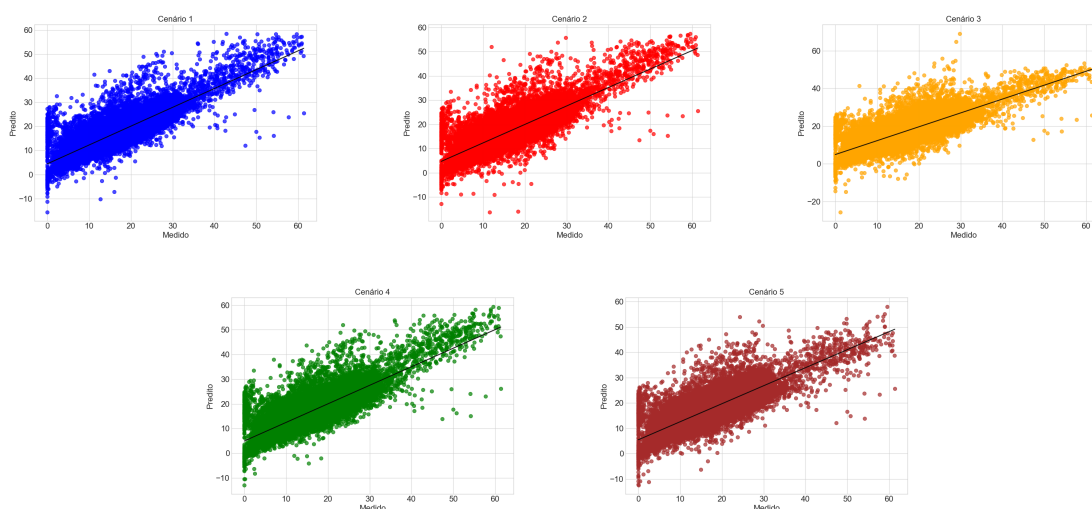


Figura 13 – Diagramas de dispersão para as melhores execuções individuais de cada estação

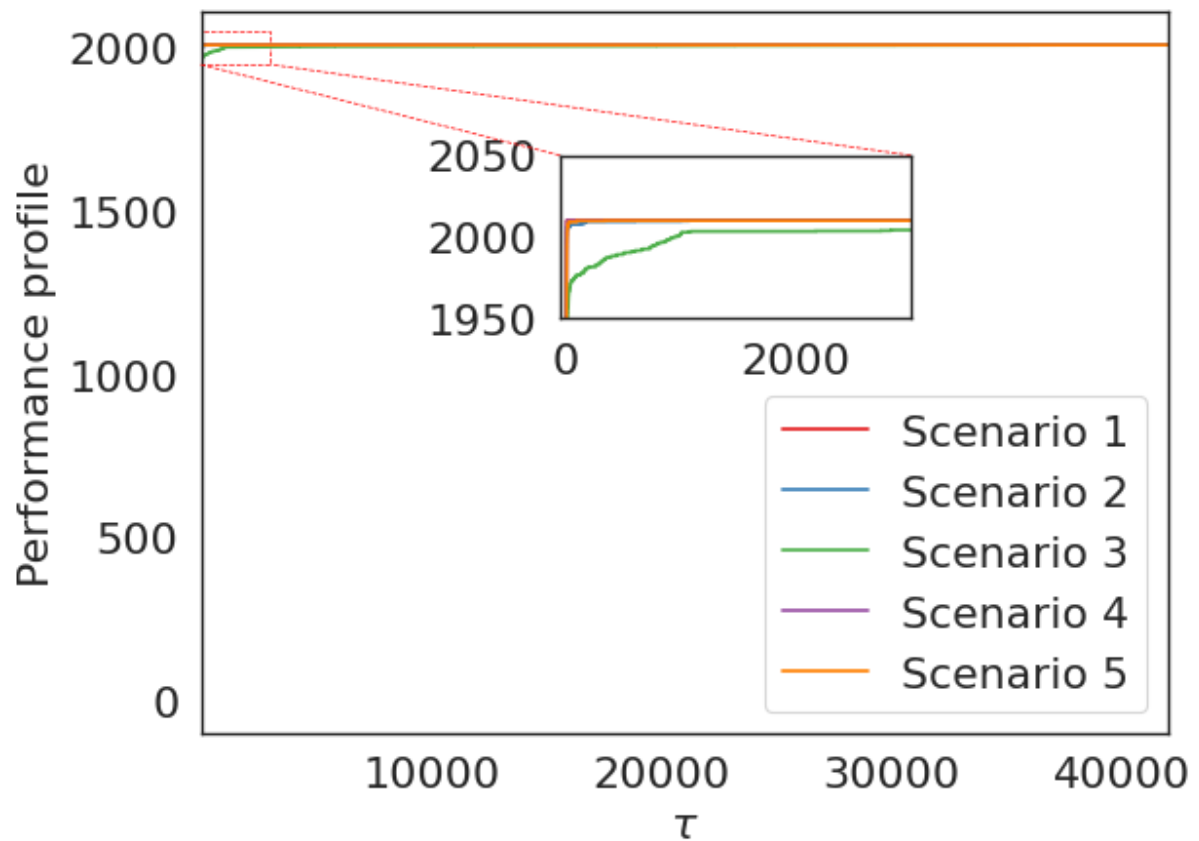


Figura 14 – Performance Profiles

Tabela 6 – Áreas normalizadas dos perfis de desempenho - Etapa 2

Cenários	Áreas Normalizadas
1	1.0
2	0.9999882749674386
3	0.9984366426408016
4	0.9999983521300365
5	0.9999879328055873

4.3 TERCEIRA ETAPA DE EXPERIMENTOS

Na terceira etapa dos experimentos buscamos entender como as cidades que compoem a base de treino influenciam os resultados. Para tal, fixamos a Estação Alvo como Juiz de Fora e utilizamos um otimizador para seleção do conjunto de estações que comporão a base de treino. Os dados da estação de Juiz de Fora foram utilizados somente na etapa de teste. Nesta etapa foram realizadas 10 execuções individuais.

Considerando esta abordagem como um sexto cenário de comparação, a Tabela 7 apresenta os valores dos seis cenários para a cidade de Juiz de Fora. A Tabela Suplementar

S4 mostra em quais execuções cada cidade foi selecionada pelo algoritmo otimizador na composição da base de treino para este experimento, assim como o somatório das utilizações e os valores de R^2 em cada execução. O mesmo resultado pode ser visto de forma gráfica na Figura 15.

Tabela 7 – Média das métricas de desempenho para a cidade de Juiz de Fora

Cenário	R^2	WI	RMSE	MAE	MBE
1	0.843	0.955	2.41	1.89	-0.051
2	0.826	0.949	2.56	2.01	-0.124
3	-0.662	0.715	7.41	6.71	-6.25
4	0.808	0.942	2.71	2.15	-0.484
5	0.619	0.893	3.76	3.09	-2.25
6	0.715	0.906	3.29	2.69	-1.30

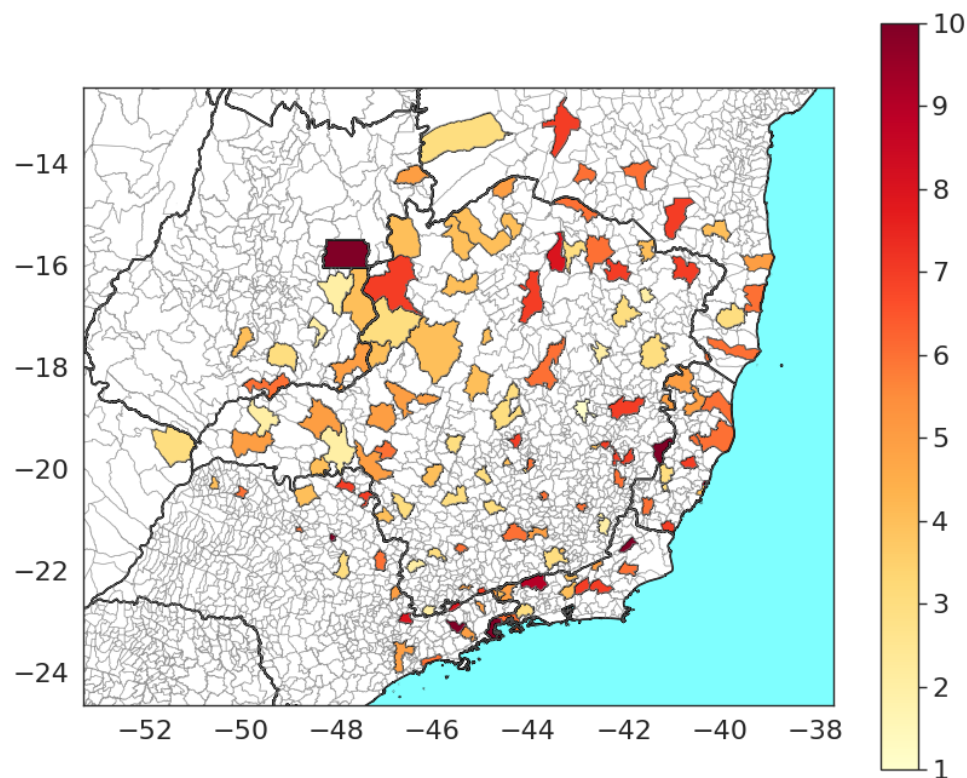


Figura 15 – *Heatmap* da utilização das cidades na formação da base de treino

4.3.1 DISCUSSÃO

Observando os resultados da Tabela 7 e comparando especificamente os cenários 3, 5 e 6, em que a técnica de transferência de aprendizado foi aplicada, vemos que utilizar o algoritmo otimizador para a seleção das cidades foi proveitoso. Apesar de os resultados ainda não alcançarem os dos cenários em que não existe transferência de aprendizado, eles foram substancialmente melhores do que selecionando as bases de treino apenas pela distância.

Quanto a utilização das cidades na formação da base de treino, cada cidade foi utilizada em média, 5.0373 vezes e cada execução utilizou em média 67.5 cidades. Todas as cidades foram selecionadas ao final de uma execução pelo menos uma vez. As cidades de Aimorés - MG, Brasília - DF, Cambuci - RJ, Paraty - RJ, Pradópolis - SP e Taubaté -SP estiveram presentes ao final de todas as 10 execuções. A falta de padrão na utilização das cidades sugere que a frequência de utilização é direcionada unicamente por quanto os dados melhor explicam a variabilidade do local de destino em demérito a outros fatores, como simples proximidade física ou características geográficas. Destacamos a utilização de Guanhães, que teve um bom resultado individual (Tabela S2), porém foi utilizada apenas uma única vez ao final de uma execução.

Considerando o resultado superior do cenário 6 sobre os cenários 3 e 5, a quantidade de cidades utilizadas na construção da base de treino deste cenário e a tendência já observada na etapa anterior de que o aumento da quantidade de cidades utilizadas na base de treino melhora o resultado final, vemos que a proposta de entregar ao algoritmo otimizador a seleção das cidades para composição da base de treino é interessante, e até mesmo necessária para a obtenção de melhores resultados. Quanto à escolha das cidades feita pelo algoritmo otimizador, buscamos encontrar algum padrão relacionado à distribuição geográfica das cidades, qualidade, tamanho e abrangência da base de dados da cidade escolhida, o resultado da cidade no cenário 1, a distância da cidade até Juiz de Fora, mas todas as análises foram inconclusivas. Esta situação reforça ainda mais a necessidade de uma seleção automática da base, visto que para uma seleção manual não encontramos uma métrica norteadora.

4.4 PONTOS FORTES E LIMITAÇÕES DO MODELO

Com esses resultados, é possível observar que, na maioria dos casos, o uso de transferência de aprendizado é uma boa opção para prever a irradiação solar em locais sem estação meteorológica, mesmo que, em algumas estações, os resultados não sejam satisfatórios. Observando a melhoria entre os resultados do cenário 5 e do cenário 3, também podemos constatar que o número de cidades utilizadas na base de treinamento é relevante, visto que, neste estudo, em 46 estações, os resultados melhoraram com o aumento do número de bases na base de treinamento. A utilização de um número ainda

maior de cidades na terceira etapa de experimentos reforça essa tendência, mostrando também a validade da escolha automática das bases de treino no processo de transferência de aprendizado.

Destacamos como limitações do modelo o alto custo computacional do processo de treinamento e a incapacidade de modelos de transferência de aprendizado em performarem melhor do que modelos treinados com as bases alvo.

5 CONCLUSÃO

Este trabalho apresenta a metodologia e os resultados da aplicação de dados de geolocalização na construção de modelos computacionais para predição da irradiação solar em Minas Gerais, Brasil. Dados de 134 cidades, adquiridos ao longo de mais de 20 anos de observações, foram analisados para responder as três questões principais:

1. A adição de dados de geolocalização para predição da irradiação solar melhora os resultados?
2. É possível prever valores de irradiação solar de forma eficiente em um local sem estações meteorológicas?
3. Como identificar o melhor conjunto de dados para treinamento de um modelo de aprendizado?

Foram realizados testes em cinco cenários diferentes, nos quais verificamos os resultados independentes em cada uma das 67 cidades de Minas Gerais que possuem estações meteorológicas, analisamos a inclusão de dados de geolocalização de cidades vizinhas para melhorar os resultados e aplicamos a transferência de aprendizado em uma metodologia que permite a previsão em cidades sem estações meteorológicas. No sexto cenário buscamos entender como a composição da base de treinamento influencia os resultados da transferência de aprendizado.

Os resultados do cenário 1, com a execução independente de cada estação, foram comparados a outros trabalhos encontrados na literatura e serviram como parâmetro de comparação para os demais estudos realizados. Para solução da questão 1, nos cenários 2 e 4, adicionamos dados de geolocalização de cidades vizinhas e **constatamos que esse tipo de abordagem geralmente piora os resultados obtidos em comparação com o cenário 1**. Nos cenários 3 e 5, utilizamos dados das estações vizinhas às cidades estudadas para compor as bases de treinamento. Utilizamos dados de cada cidade analisada como base de teste, aplicando, assim, a técnica de transferência de aprendizado. **O estudo desses cenários demonstrou que, na maioria dos casos, é possível prever com eficiência os valores de irradiação solar mesmo quando não há estações meteorológicas, utilizando os bancos de dados de estações em cidades vizinhas**, respondendo à questão 2. O estudo desses cenários também demonstra que a escolha do número de cidades utilizadas no processo de criação do banco de dados de treinamento influencia os resultados obtidos. No cenário 6, encontramos a resposta à questão 3, quando **delegamos ao otimizador a tarefa de selecionar as melhores cidades para composição da base de treino** em um estudo de caso focado na cidade de Juiz de Fora.

A principal conquista deste trabalho é o desenvolvimento de modelos para prever a irradiação solar em cidades sem estações meteorológicas. O grande número de cidades

analisadas permitiu-nos verificar que as técnicas de transferência de aprendizado funcionam para a maioria dos locais. As contribuições deste trabalho são resumidas a seguir:

1. Uma metodologia para transferência de aprendizado para locais geograficamente dispersos;
2. Metodologia para o desenvolvimento de um modelo computacional otimizado;
3. Seleção automática de recursos;
4. Aplicação inédita do modelo ELM para a previsão da irradiação solar em Minas Gerais;
5. Estudo abrangendo todo o estado de Minas Gerais;
6. Seleção automática da base de treino para o processo de transferência de aprendizado.

Este trabalho representa um avanço significativo na previsão da irradiação solar em regiões com carência de estações meteorológicas, particularmente em Minas Gerais, Brasil. A metodologia proposta, que combina dados de geolocalização e técnicas de transferência de aprendizado, demonstrou ser eficaz na obtenção de alta precisão nas previsões de irradiação solar.

A seleção inteligente de estações próximas pode impulsionar melhorias no modelo para compor o banco de dados de treinamento. A melhor escolha de estações pode impactar significativamente o desempenho do modelo, e pesquisas futuras devem se concentrar em expandir os locais analisados para generalizar os resultados encontrados.

Apesar das limitações, este estudo contribui com metodologias e percepções valiosos para propor previsões de irradiação solar mais precisas e para o avanço do planejamento e implementação de energias renováveis. Os resultados deste estudo podem ser utilizados para desenvolver sistemas de energia solar mais eficientes e econômicos, o que pode ajudar a reduzir nossa dependência de combustíveis fósseis e combater as mudanças climáticas.

5.1 POSSÍVEIS DESDOBRAMENTOS

- Aplicação da metodologia da terceira etapa em um conjunto maior de cidades;
- Estudo comparativo dos resultados encontrados utilizando os dados do INMET com dados de satélite;
- Aplicação da transferência de aprendizado utilizando como base de teste dados provenientes de outra fonte que não seja o INMET;
- Aprofundamento das questões relativas à geração fotovoltaica na rede de transmissão regional;
- Desenvolvimento de um modelo online.

REFERÊNCIAS

- 1 Pysolar: staring directly at the sun since 2007, 2008. URL <http://pysolar.readthedocs.io/>.
- 2 Agência Nacional de Energia Elétrica. Sistema de Informações de Geração da ANEEL, 2025. URL <https://app.powerbi.com/view?r=eyJrIjoiY2VmMmUwN2QtYWFiOS00ZDE3LWI3NDMtZDk0NGI4MGU2NTkxIiwidCI6IjQwZDZmOWI4LWVjYTctNDZhMi05MmQ0LWVhNGU5YzAxNzBlMSIsImMiOjR9>.
- 3 Muhammad Naveed Akhter, Saad Mekhilef, Hazlie Mokhlis, and Noraisyah Mohamed Shah. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renewable Power Generation*, 13(7):1009–1023, 2019. doi: <https://doi.org/10.1049/iet-rpg.2018.5649>. URL <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2018.5649>.
- 4 J. Antonanzas, D. Pozo-Vázquez, L.A. Fernandez-Jimenez, and F.J. Martinez de Pison. The value of day-ahead forecasting for photovoltaics in the spanish electricity market. *Solar Energy*, 158:140–146, 2017. ISSN 0038-092X. doi: <https://doi.org/10.1016/j.solener.2017.09.043>. URL <https://www.sciencedirect.com/science/article/pii/S0038092X17308307>.
- 5 Associação Brasileira de Energia Solar Fotovoltaica. Energia Solar Fotovoltaica no Brasil, 2025. URL <https://www.absolar.org.br>.
- 6 Nasrin Fathollahzadeh Attar, Mohammad Taghi Sattari, Ramendra Prasad, and Halit Apaydin. Comprehensive review of solar radiation modeling based on artificial intelligence and optimization techniques: future concerns and considerations. *Clean Technologies and Environmental Policy*, 25(4):1079–1097, May 2023. ISSN 1618-9558. doi: [10.1007/s10098-022-02434-7](https://doi.org/10.1007/s10098-022-02434-7). URL <https://doi.org/10.1007/s10098-022-02434-7>.
- 7 O.M. Babatunde, J.L. Munda, Y. Hamam, and C.G. Monyei. A critical overview of the (im)practicability of solar radiation forecasting models. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 5:100213, 2023. ISSN 2772-6711. doi: <https://doi.org/10.1016/j.prime.2023.100213>. URL <https://www.sciencedirect.com/science/article/pii/S2772671123001080>.
- 8 Lucas A. Barbosa, Cláudio R. Silva, Marco Túlio G. Paula, Valdiney J. Silva, Rafael R. Finzi, Fábio J. Carvalho, and Heyder F. Nascimento. Estimativa da radiação solar com base na temperatura do ar na região sul, suldeste, oeste de Minas e Campo das Vertentes. In *XVII Congresso Brasileiro de Agrometeorologia, 2011*, Guarapari - ES, 2011. URL <http://www.sbagro.org/files/biblioteca/3194.pdf>.
- 9 Samuel Basilio and Leonardo Goliatt. Gradient boosting hybridized with exponential natural evolution strategies for estimating the strength of geopolymer self-compacting concrete. *Knowledge-Based Engineering and Sciences*, 3:1–16, 04 2022. doi: [10.51526/kbes.2022.3.1.1-16](https://doi.org/10.51526/kbes.2022.3.1.1-16).

- 10 Samuel Basilio, , and Leonardo Goliatt. An automl framework for transfer learning applied to solar radiation prediction: A case study in juiz de fora, minas gerais, brazil. *XLVI CILAMCE*, 11 2025. doi: <https://doi.org/10.55592/cilamce2025.v5i.13251>.
- 11 Samuel da Costa Alves Basílio, Camila Martins Saporette, Zaher Mundher Yaseen, and Leonardo Goliatt. Global horizontal irradiance modeling from environmental inputs using machine learning with automatic model selection: A case study in tanzania. *Environmental Development*, page 100766, 2022. ISSN 2211-4645. doi: <https://doi.org/10.1016/j.envdev.2022.100766>. URL <https://www.sciencedirect.com/science/article/pii/S2211464522000689>.
- 12 Samuel da Costa Alves Basílio, Rodrigo O. Silva, Camila M. Saporette, and Leonardo Goliatt. Modeling global solar radiation using machine learning with model selection approach: A case study in tanzania. In Subarna Shakya, Klimis Ntalianis, and Khaled A. Kamel, editors, *Mobile Computing and Sustainable Informatics*, pages 155–168, Singapore, 2022. Springer Nature Singapore. ISBN 978-981-19-2069-1.
- 13 Samuel da Costa Alves Basílio, Fernando Ferrari Putti, Angélica Carvalho Cunha, and Leonardo Goliatt. An evolutionary-assisted machine learning model for global solar radiation prediction in Minas Gerais region, southeastern Brazil. *Earth Science Informatics*, Apr 2023. ISSN 1865-0481. doi: 10.1007/s12145-023-00990-0. URL <https://doi.org/10.1007/s12145-023-00990-0>.
- 14 Samuel da Costa Alves Basílio, Camila M. Saporette, and Leonardo Goliatt. An interdependent evolutionary machine learning model applied to global horizontal irradiance modeling. *Neural Computing and Applications*, 35(16):12099–12120, Jun 2023. ISSN 1433-3058. doi: 10.1007/s00521-023-08342-1. URL <https://doi.org/10.1007/s00521-023-08342-1>.
- 15 Ricardo J. Bessa, Artur Trindade, and Vladimiro Miranda. Spatial-temporal solar power forecasting for smart grids. *IEEE Transactions on Industrial Informatics*, 11(1):232–241, 2015. doi: 10.1109/TII.2014.2365703.
- 16 Kelin Pribs Bexaira, Nereu Augusto Streck, Jossana Ceolin Cera, and Solismar Damé Prestes. Coeficientes de angström-prescott para estimar a radiação solar no rio grande do sul. *Revista Brasileira de Meteorologia*, 33(3):401–411, 2018. ISSN 1982-4351.
- 17 Fatih Birol. World Energy Outlook 2021. Technical report, International Energy Agency, 2021. URL <https://www.iea.org/reports/world-energy-outlook-2021>.
- 18 Francesco Biscani and Dario Izzo. A parallel global multiobjective framework for optimization: pagmo. *Journal of Open Source Software*, 5(53):2338, 2020.
- 19 Tales Boratto, Camila Saporette, Samuel Basilio, Alexandre Cury, and Leonardo Goliatt. Data-driven cymbal bronze alloy identification via evolutionary machine learning with automatic feature selection. *Journal of Intelligent Manufacturing*, 35: 1–17, 11 2022. doi: 10.1007/s10845-022-02047-3.
- 20 Brasil. PL 1917/2015, 2015. URL <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=1307190>.

- 21 Brasil. Portaria nº 465, de 12 de dezembro de 2019, 2019. URL <https://www.in.gov.br/en/web/dou/-/portaria-n-465-de-12-de-dezembro-de-2019.-233554889>.
- 22 Brasil. PL 414/2021, 2021. URL <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2270036>.
- 23 Tatiane C. Carneiro, Paulo A.C. Rocha, Paulo C.M. Carvalho, and Luis M. Fernández-Ramírez. Ridge regression ensemble of machine learning models applied to solar and wind forecasting in brazil and spain. *Applied Energy*, 314:118936, 2022. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2022.118936>. URL <https://www.sciencedirect.com/science/article/pii/S0306261922003555>.
- 24 Tatiane Carolyne Carneiro, Paulo Cesar Marques de Carvalho, Heron Alves dos Santos, Marcello Anderson Ferreira Batista Lima, and Arthur Plinio de Souza Braga. Review on Photovoltaic Power and Solar Resource Forecasting: Current Status and Trends. *Journal of Solar Energy Engineering*, 144(1), 07 2021. ISSN 0199-6231. doi: 10.1115/1.4051652. URL <https://doi.org/10.1115/1.4051652>. 010801.
- 25 Fábio J. Carvalho, Cláudio R. Silva, Heyder F. Nascimento, Lucas A. Barbosa, Marco Túlio G. Paula, Valdiney J. Silva, and Rafael R. Finzi. Avaliação de modelos de estimativa da radiação solar com base na temperatura do ar para o norte de Minas Gerais. In *XVII Congresso Brasileiro de Agrometeorologia, 2011*, Guarapari - ES, 2011. URL <http://www.sbagro.org/files/biblioteca/3179.pdf>.
- 26 Micaele Martins de Carvalho, Aline Souza Magalhães, and Edson Paulo Domingues. Impactos econômicos da ampliação do uso de energia solar residencial em minas gerais. *Nova economia (Belo Horizonte, Brazil)*, 29(2):459, 2019. ISSN 0103-6351.
- 27 João Rodrigo de Castro, Santiago Vianna Cuadra, Luciana Barros Pinto, João Marcelo Hoffmann de Souza, Marcos Paulo dos Santos, and Alexandre Bryan Heinemann. Parametrization of models and use of estimated global solar radiation data in the irrigated rice yield simulation. *Revista Brasileira de Meteorologia*, 33(2): 238–246, 2018. ISSN 0102-7786.
- 28 Songjian Chai, Zhao Xu, Youwei Jia, and Wai Kin Wong. A robust spatiotemporal forecasting framework for photovoltaic generation. *IEEE Transactions on Smart Grid*, 11(6):5370–5382, 2020. doi: 10.1109/TSG.2020.3006085.
- 29 Tianqi Chen and Carlos Guestrin. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, August 2016. doi: 10.1145/2939672.2939785. URL <https://doi.org/10.1145/2939672.2939785>.
- 30 Min-Yuan Cheng and Minh-Tu Cao. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Applied Soft Computing*, 22:178 – 188, 2014. ISSN 1568-4946.
- 31 Enrique Chueca, Mariana Weiss, Rogelio Celaya, Pauline Ravillard, Boris Ortega, Mauricio T. Tolmasquim, and Michelle Hallack. Early adopters of residential solar pv distributed generation: Evidence from brazil, chile and mexico. *Energy for Sustainable Development*, 76:101284, 2023. ISSN 0973-0826. doi: <https://doi.org/10.1016/j.esd.2023.101284>. URL <https://www.sciencedirect.com/science/article/pii/S0973082623001412>.

- 32 Jenny Cifuentes, Geovanny Marulanda, Antonio Bello, and Javier Reneses. Air temperature forecasting using machine learning techniques: A review. *Energies*, 13(16), 2020. ISSN 1996-1073. doi: 10.3390/en13164215. URL <https://www.mdpi.com/1996-1073/13/16/4215>.
- 33 Angélica Carvalho Cunha, Luís Roberto Almeida Gabriel Filho, Adriana Aki Tanaka, and Fernando Ferrari Putti. Performance and estimation of solar radiation models in state of minas gerais, brazil. *Modeling Earth Systems and Environment*, 7(1):603–622, Mar 2021. ISSN 2363-6211. doi: 10.1007/s40808-020-00956-x. URL <https://doi.org/10.1007/s40808-020-00956-x>.
- 34 Samuel da Costa Alves Basílio, Cicero Manoel dos Santos, Alfeu Dias Martinho, and Leonardo Goliatt. Transfer learning for solar irradiation prediction in minas gerais, brazil. *Scientific Reports*, 15(1):40235, Nov 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-24095-4. URL <https://doi.org/10.1038/s41598-025-24095-4>.
- 35 Maurício Bruno Prado da Silva, João Francisco Escobedo, Taiza Juliana Rossi, Cícero Manoel dos Santos, and Sílvia Helena Modenese Gorla da Silva. 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*, 160:11–23, 2017. ISSN 1364-6826. doi: <https://doi.org/10.1016/j.jastp.2017.04.001>. URL <https://www.sciencedirect.com/science/article/pii/S1364682616304448>.
- 36 Antonio Augusto Aguiar Dantas, Luiz Gonsaga de Carvalho, and Elizabeth Ferreira. Estimativa da radiação solar global para a região de lavras, mg. *Ciência e Agrotecnologia*, 27:1260–1263, 2003.
- 37 Utpal Kumar Das, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh Yamani Idna Idris, Willem Van Deventer, Bend Horan, and Alex Stojcevski. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81:912–928, 2018. ISSN 1364-0321. doi: <https://doi.org/10.1016/j.rser.2017.08.017>. URL <https://www.sciencedirect.com/science/article/pii/S1364032117311620>.
- 38 Gabriel de Freitas Viscondi and Solange N. Alves-Souza. A systematic literature review on big data for solar photovoltaic electricity generation forecasting. *Sustainable Energy Technologies and Assessments*, 31:54–63, 2019. ISSN 2213-1388. doi: <https://doi.org/10.1016/j.seta.2018.11.008>. URL <https://www.sciencedirect.com/science/article/pii/S2213138818301036>.
- 39 Gabriel de Freitas Viscondi and Solange N. Alves-Souza. Solar irradiance prediction with machine learning algorithms: A brazilian case study on photovoltaic electricity generation. *Energies*, 14(18), 2021. ISSN 1996-1073. URL <https://www.mdpi.com/1996-1073/14/18/5657>.
- 40 Domingos S. de O. Santos, Paulo S. G. de Mattos Neto, João F. L. de Oliveira, Hugo Valadares Siqueira, Tathiana Mikamura Barchi, Aranildo R. Lima, Francisco Madeiro, Douglas A. P. Dantas, Attilio Converti, Alex C. Pereira, José Bione de Melo Filho, and Manoel H. N. Marinho. Solar irradiance forecasting using

- dynamic ensemble selection. *Applied Sciences*, 12(7), 2022. ISSN 2076-3417. doi: 10.3390/app12073510. URL <https://www.mdpi.com/2076-3417/12/7/3510>.
- 41 Iara F. B. De Oliveira, Leandro A. Vasconcelos, and Igor S. O. Freitas. Impact of dirt on the performance of photovoltaic plants in the north of minas gerais. In *2023 15th IEEE International Conference on Industry Applications (INDUSCON)*, pages 858–863, 2023. doi: 10.1109/INDUSCON58041.2023.10374784.
- 42 Vahdettin Demir and Hatice Citakoglu. Forecasting of solar radiation using different machine learning approaches. *Neural Computing and Applications*, 35(1):887–906, Jan 2023. ISSN 1433-3058. doi: 10.1007/s00521-022-07841-x. URL <https://doi.org/10.1007/s00521-022-07841-x>.
- 43 Elizabeth D. Dolan and Jorge J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, Jan 2002. ISSN 1436-4646. doi: 10.1007/s101070100263. URL <https://doi.org/10.1007/s101070100263>.
- 44 Cícero Manoel dos Santos, Érico Tadao Teramoto, Amaury de Souza, Flavio Aristone, and Razika Ihaddadene. Several models to estimate daily global solar irradiation: adjustment and evaluation. *Arabian Journal of Geosciences*, 14(4):286, Feb 2021. ISSN 1866-7538. doi: 10.1007/s12517-021-06603-8. URL <https://doi.org/10.1007/s12517-021-06603-8>.
- 45 Cícero Manoel dos Santos, João Francisco Escobedo, Érico Tadao Teramoto, and Silvia Helena Modenese Gorla da Silva. Assessment of ann and svm models for estimating normal direct irradiation (hb). *Energy Conversion and Management*, 126: 826–836, 2016. ISSN 0196-8904. doi: <https://doi.org/10.1016/j.enconman.2016.08.020>. URL <https://www.sciencedirect.com/science/article/pii/S019689041630694X>.
- 46 Boudewijn Elsinga and Wilfried G.J.H.M. van Sark. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. *Applied Energy*, 206:1464–1483, 2017. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2017.09.115>. URL <https://www.sciencedirect.com/science/article/pii/S0306261917314010>.
- 47 Diego Simões Fernandes, Alexandre Bryan Heinemann, André de Oliveira Amorim, and Rosidalva Lopes Feitosa da Paz. Estimativa da Radiação Solar Global com Base em Observações de Temperatura para o Estado de Goiás. *Revista Brasileira de Meteorologia*, 33:558 – 566, 09 2018. ISSN 0102-7786. URL http://old.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862018000300558&nrm=iso.
- 48 Lucas Borges Ferreira, Fernando França da Cunha, Rubens Alves de Oliveira, and Elpídio Inácio Fernandes Filho. Estimation of reference evapotranspiration in brazil with limited meteorological data using ann and svm – a new approach. *Journal of Hydrology*, 572:556–570, 2019. ISSN 0022-1694. doi: <https://doi.org/10.1016/j.jhydrol.2019.03.028>. URL <https://www.sciencedirect.com/science/article/pii/S0022169419302689>.
- 49 Lucas Borges Ferreira, Fernando França da Cunha, and Elpídio Inácio Fernandes Filho. Exploring machine learning and multi-task learning to estimate meteorological data and reference evapotranspiration across brazil. *Agricultural Water Management*,

- 259:107281, 2022. ISSN 0378-3774. doi: <https://doi.org/10.1016/j.agwat.2021.107281>. URL <https://www.sciencedirect.com/science/article/pii/S0378377421005588>.
- 50 Lucas Santos do Patrocínio Figueiró, Caroline Salezzi Bonfá, and Lucas da Costa Santos. Calibration and evaluation of the Hargreaves-Samani equation for estimating reference evapotranspiration: a case study in northern minas gerais, brazil. *Rev. Gest. Secr.*, 16(2):e4638, February 2025.
- 51 Rafael R. Finzi, Cláudio R. Silva, Fábio J. Carvalho, Heyder F. Nascimento, Lucas A. Barbosa, Marco Túlio G. Paula, and Valdiney J. Silva. Estimativa da radiação solar baseando-se na temperatura máxima e mínima do ar para a região noroeste de Minas Gerais. In *XVII Congresso Brasileiro de Agrometeorologia, 2011*, Guarapari - ES, 2011. URL <http://www.sbagro.org/files/biblioteca/3176.pdf>.
- 52 Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1–22, 2010. ISSN 1548-7660.
- 53 Jerome H Friedman. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.
- 54 Ricardo H.G. Furiati, Filipe Sacchetto, Simon Malinowski, Zenilton Kleber G. do Patrocínio, Felipe D. Cunha, Cristiana B. Maia, and Silvio Jamil F. Guimarães. Solar radiation times-series forecasting in southern brazil: A comprehensive analysis. *Ecological Informatics*, page 103601, 2026. ISSN 1574-9541. doi: <https://doi.org/10.1016/j.ecoinf.2026.103601>. URL <https://www.sciencedirect.com/science/article/pii/S1574954126000075>.
- 55 Chaitanya Ganvir, D. Dinesh, Rahul Gupta, SK Jha, and Praveen Kumar Raghuvanshi. Prediction of global horizontal irradiance based on explainable artificial intelligence. In *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, pages 1–4, 2024. doi: 10.1109/IITCEE59897.2024.10467440.
- 56 Sujan Ghimire, Ravinesh C. Deo, Nathan J. Downs, and Nawin Raj. Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed modis satellite and reanalysis atmospheric products in solar-rich cities. *Remote Sensing of Environment*, 212:176–198, 2018. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2018.05.003>. URL <https://www.sciencedirect.com/science/article/pii/S0034425718302165>.
- 57 Mawloud Guermoui, Farid Melgani, Kacem Gairaa, and Mohamed Lamine Mekhalfi. A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production*, 258:120357, 2020. ISSN 0959-6526. doi: <https://doi.org/10.1016/j.jclepro.2020.120357>. URL <https://www.sciencedirect.com/science/article/pii/S0959652620304042>.
- 58 Rahul Gupta, Anil Kumar Yadav, SK Jha, and Pawan Kumar Pathak. Comparative analysis of advanced machine learning classifiers based on feature engineering framework for weather prediction. *Scientia Iranica*, pages –, 2024. ISSN 1026-3098. doi: 10.24200/sci.2024.61305.7242. URL https://scientiairanica.sharif.edu/article_23690.html.

- 59 Rahul Gupta, Anil Kumar Yadav, S.K. Jha, and Pawan Kumar Pathak. Composition of feature selection techniques for improving the global horizontal irradiance estimation via machine learning models. *Thermal Science and Engineering Progress*, 48:102394, 2024. ISSN 2451-9049. doi: <https://doi.org/10.1016/j.tsep.2024.102394>. URL <https://www.sciencedirect.com/science/article/pii/S245190492400012X>.
- 60 Ali Etem Gürel, Ümit Ağbulut, Hüseyin Bakır, Alper Ergün, and Gökhan Yıldız. A state of art review on estimation of solar radiation with various models. *Heliyon*, 9 (2):e13167, February 2023.
- 61 Federico-Vladimir Gutierrez-Corea, Miguel-Angel Manso-Callejo, Maria-Pilar Moreno-Regidor, and Maria-Teresa Manrique-Sancho. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Solar Energy*, 134:119–131, 2016. ISSN 0038-092X. doi: <https://doi.org/10.1016/j.solener.2016.04.020>. URL <https://www.sciencedirect.com/science/article/pii/S0038092X16300536>.
- 62 Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501, 2006. ISSN 0925-2312. doi: <https://doi.org/10.1016/j.neucom.2005.12.126>. URL <https://www.sciencedirect.com/science/article/pii/S0925231206000385>. Neural Networks.
- 63 Ahmedbahaaaldin Ibrahim Ahmed Osman, Ali Najah Ahmed, Ming Fai Chow, Yuk Feng Huang, and Ahmed El-Shafie. Extreme gradient boosting (xgboost) model to predict the groundwater levels in selangor malaysia. *Ain Shams Engineering Journal*, 12(2):1545–1556, 2021. ISSN 2090-4479.
- 64 Instituto Brasileiro de Geografia e Estatística. *Biomass do Brasil*, 2019. URL <http://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15842-biomass.html>. Available at <http://www.ibge.gov.br>.
- 65 Instituto Nacional de Meteorologia, 2025. URL <http://portal.inmet.gov.br/>.
- 66 Ankit Jha, Vansh Goel, Manish Kumar, Gulshan Kumar, Rahul Gupta, and Shyama Kant Jha. An efficient and interpretable stacked model for wind speed estimation based on ensemble learning algorithms. *Energy Technology*, 12(6):2301188, 2024. doi: <https://doi.org/10.1002/ente.202301188>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/ente.202301188>.
- 67 Chengcheng Jiang and Qunzhi Zhu. Evaluating the most significant input parameters for forecasting global solar radiation of different sequences based on informer. *Applied Energy*, 348:121544, 2023. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2023.121544>. URL <https://www.sciencedirect.com/science/article/pii/S030626192300908X>.
- 68 F. V. Carvalho Júnior, M. C. Alves, F. S. Menezes, and L. G. Carvalho. Multiple effect of climatic factors and elements on photovoltaic generation in on-grid system using conventional station data. *Applied Solar Energy*, Nov 2025. ISSN 1934-9424. doi: [10.3103/S0003701X25601036](https://doi.org/10.3103/S0003701X25601036). URL <https://doi.org/10.3103/S0003701X25601036>.

- 69 A. Khosravi, L. Machado, and R.O. Nunes. Time-series prediction of wind speed using machine learning algorithms: A case study osorio wind farm, brazil. *Applied Energy*, 224:550–566, 2018. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2018.05.043>. URL <https://www.sciencedirect.com/science/article/pii/S0306261918307360>.
- 70 Naveen Krishnan, K. Ravi Kumar, and Chandrapal Singh Inda. How solar radiation forecasting impacts the utilization of solar energy: A critical review. *Journal of Cleaner Production*, 388:135860, 2023. ISSN 0959-6526. doi: <https://doi.org/10.1016/j.jclepro.2023.135860>. URL <https://www.sciencedirect.com/science/article/pii/S0959652623000185>.
- 71 Pratima Kumari and Durga Toshniwal. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Applied Energy*, 295:117061, 2021. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2021.117061>. URL <https://www.sciencedirect.com/science/article/pii/S0306261921005158>.
- 72 Hai Lan, He Yin, Ying-Yi Hong, Shuli Wen, David C. Yu, and Peng Cheng. Day-ahead spatio-temporal forecasting of solar irradiation along a navigation route. *Applied Energy*, 211:15–27, 2018. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2017.11.014>. URL <https://www.sciencedirect.com/science/article/pii/S0306261917315945>.
- 73 Francisco J.L. Lima, Fernando R. Martins, Enio B. Pereira, Elke Lorenz, and Detlev Heinemann. Forecast for surface solar irradiance at the brazilian northeastern region using nwp model and artificial neural networks. *Renewable Energy*, 87:807–818, 2016. ISSN 0960-1481. doi: <https://doi.org/10.1016/j.renene.2015.11.005>. URL <https://www.sciencedirect.com/science/article/pii/S0960148115304249>.
- 74 Marcello Anderson F.B. Lima, Paulo C.M. Carvalho, Luis M. Fernández-Ramírez, and Arthur P.S. Braga. Improving solar forecasting using deep learning and portfolio theory integration. *Energy*, 195:117016, 2020. ISSN 0360-5442. doi: <https://doi.org/10.1016/j.energy.2020.117016>. URL <https://www.sciencedirect.com/science/article/pii/S0360544220301237>.
- 75 Guanjun Liu, Hui Qin, Qin Shen, Hao Lyv, Yuhua Qu, Jialong Fu, Yongqi Liu, and Jianzhong Zhou. Probabilistic spatiotemporal solar irradiation forecasting using deep ensembles convolutional shared weight long short-term memory network. *Applied Energy*, 300:117379, 2021. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2021.117379>. URL <https://www.sciencedirect.com/science/article/pii/S0306261921007820>.
- 76 Yongqi Liu, Hui Qin, Zhendong Zhang, Shaoqian Pei, Chao Wang, Xiang Yu, Zhiqiang Jiang, and Jianzhong Zhou. Ensemble spatiotemporal forecasting of solar irradiation using variational bayesian convolutional gate recurrent unit network. *Applied Energy*, 253:113596, 2019. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2019.113596>. URL <https://www.sciencedirect.com/science/article/pii/S030626191931270X>.
- 77 Gustavo Bastos Lyra, Sidney Sára Zanetti, Anderson Amorim Rocha Santos, José Leonardo de Souza, Guilherme Bastos Lyra, José Francisco Oliveira-Júnior, and

- Marco Antônio Maringolo Lemes. Estimation of monthly global solar irradiation using the Hargreaves–Samani model and an artificial neural network for the state of Alagoas in northeastern Brazil. *Theoretical and applied climatology*, 125(3-4):743–756, 2015. ISSN 0177-798X.
- 78 Gustavo Bastos Lyra, Cicero Manoel dos Santos, Marcel Carvalho Abreu, Allana Oliveira Lima, José Leonaldo de Souza, Guilherme Bastos Lyra, Sidney Sára Zanetti, and Fabrina Bolzan Martins. Estimates of monthly global solar irradiation using empirical models and artificial intelligence techniques based on air temperature in southeastern Brazil. *Theoretical and Applied Climatology*, 152(3):1031–1051, May 2023. ISSN 1434-4483. doi: 10.1007/s00704-023-04442-z. URL <https://doi.org/10.1007/s00704-023-04442-z>.
- 79 Ministério de Minas e Energia. Potencial dos Recursos Energéticos no Horizonte 2050. Technical report, Empresa de Pesquisa Energética, 2018. URL https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-227/topico-416/NT04%20PR_RecursosEnergeticos%202050.pdf.
- 80 Ministério de Minas e Energia. Balanço Energético Nacional. Technical report, Empresa de Pesquisa Energética, 2022. URL <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>.
- 81 Ana Flávia Martins Monteiro and Fabrina Bolzan Martins. Global Solar Radiation Models in Minas Gerais, Southeastern Brazil. *Advances in Meteorology*, 2019:9515430, Dec 2019. ISSN 1687-9309. doi: 10.1155/2019/9515430. URL <https://doi.org/10.1155/2019/9515430>.
- 82 Rafael Aldighieri Moraes and Wezer Lismar Miranda. Avaliação dos dados decendiais de precipitação, temperatura média, máxima e mínima do ar, radiação solar e evapotranspiração de referência simulados pelo modelo ecmwf para Minas Gerais. In *XVIII Congresso Brasileiro de Agrometeorologia, 2013*, Belém - PA, 2013. URL <http://www.sbagro.org/files/biblioteca/2855.pdf>.
- 83 Iulo Pessotti Moro, Roberto Avelino Cecílio, Sidney Sara Zanetti, Gustavo Bastos Lyra, Francielle Rodrigues de Oliveira, Lunalda Aparecida Vaz Pola, and José Eduardo Macedo Pezzopane. Performance of daily reference evapotranspiration estimation by different methods across Brazilian climates. *Rev. Bras. Meteorol.*, 40 (e40250027), 2025.
- 84 Faisal Nawab, Ag Sufiyan Abd Hamid, Adnan Ibrahim, Kamaruzzaman Sopian, Ahmad Fazlizan, and Mohd Faizal Fauzan. Solar irradiation prediction using empirical and artificial intelligence methods: A comparative review. *Heliyon*, 9(6): e17038, 2023. ISSN 2405-8440. doi: <https://doi.org/10.1016/j.heliyon.2023.e17038>. URL <https://www.sciencedirect.com/science/article/pii/S2405844023042469>.
- 85 Organização das Nações Unidas, 2022. URL <https://brasil.un.org/pt-br/sdgs>.
- 86 Marco Túlio G. Paula, Cláudio R. Silva, Valdiney J. Silva, Rafael R. Finzi, Fabio J. Carvalho, Heyder F. Nascimento, and Lucas A. Barboba. Estimativa da radiação solar através dos valores de temperatura registrados na região nordeste de Minas Gerais. In *XVII Congresso Brasileiro de Agrometeorologia, 2011*, Guarapari - ES, 2011. URL <http://www.sbagro.org/files/biblioteca/3183.pdf>.

- 87 Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.
- 88 Maneesha Perera, Julian De Hoog, Kasun Bandara, and Saman Halgamuge. Multi-resolution, multi-horizon distributed solar pv power forecasting with forecast combinations. *Expert Systems with Applications*, 205:117690, 2022. ISSN 0957-4174. doi: <https://doi.org/10.1016/j.eswa.2022.117690>. URL <https://www.sciencedirect.com/science/article/pii/S0957417422009836>.
- 89 Marco Pierro, David Moser, Richard Perez, and Cristina Cornaro. The value of pv power forecast and the paradox of the “single pricing” scheme: The italian case study. *Energies*, 13(15), 2020. ISSN 1996-1073. doi: 10.3390/en13153945. URL <https://www.mdpi.com/1996-1073/13/15/3945>.
- 90 João Vitor de Nóvoa Pinto, Hildo Giuseppe Garcia Caldas Nunes, Daniely Florencia Silva de Souza, Deborah Luciany Pires Costa, and Paulo Jorge de Oliveira Ponte de Souza. Calibration of Methods to Estimate Solar Irradiance in Northeastern Pará. *Revista Brasileira de Meteorologia*, 35:345 – 355, 06 2020. ISSN 0102-7786. URL http://old.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862020000200345&nrm=iso.
- 91 João Paulo Augusto Ramos, Murilo Dos Santos Vianna, and Fábio Ricardo Marin. Estimativa da radiação solar global baseada na amplitude térmica para o brasil. *Agrometeoros*, 26(1), December 2018. doi: 10.31062/agrom.v26i1.26299. URL <https://doi.org/10.31062/agrom.v26i1.26299>.
- 92 Fermín Rodríguez, Fernando Martín, Luis Fontán, and Ainhoa Galarza. Ensemble of machine learning and spatiotemporal parameters to forecast very short-term solar irradiation to compute photovoltaic generators’ output power. *Energy*, 229:120647, 2021. ISSN 0360-5442. doi: <https://doi.org/10.1016/j.energy.2021.120647>. URL <https://www.sciencedirect.com/science/article/pii/S0360544221008963>.
- 93 Juan Manuel Roldán-Fernández, Manuel Burgos-Payán, and Jesús Manuel Riquelme-Santos. Impact of domestic pv systems in the day-ahead iberian electricity market. *Solar Energy*, 217:15–24, 2021. ISSN 0038-092X. doi: <https://doi.org/10.1016/j.solener.2021.01.065>. URL <https://www.sciencedirect.com/science/article/pii/S0038092X21000918>.
- 94 Marshall Victor Chagas Santos, André Luiz de Carvalho, José Leonaldo de Souza, Mauricio Bruno Prado da Silva, Rui Palmeira Medeiros, Ricardo Araújo Ferreira Junior, Gustavo Bastos Lyra, Iêdo Teodoro, Guilherme Bastos Lyra, and Marco Antonio Maringolo Lemes. A modelling assessment of the maize crop growth, yield and soil water dynamics in the northeast of brazil. *Australian journal of crop science*, (14(6):2020):897–904, 2020. ISSN 1835-2693.
- 95 Pietros André Balbino dos Santos, Felipe Schwerz, Luiz Gonsaga de Carvalho, Victor Buono da Silva Baptista, Diego Bedin Marin, Gabriel Araújo e Silva Ferraz, Giuseppe Rossi, Leonardo Conti, and Gianluca Bambi. Machine learning and conventional methods for reference evapotranspiration estimation using

- limited-climatic-data scenarios. *Agronomy*, 13(9), 2023. ISSN 2073-4395. doi: 10.3390/agronomy13092366. URL <https://www.mdpi.com/2073-4395/13/9/2366>.
- 96 Cláudio R da Silva, Valdiney J da Silva, José Alves Júnior, and Hudson de P Carvalho. Radiação solar estimada com base na temperatura do ar para três regiões de minas gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 16:281–288, 2012.
- 97 Valdiney José da Silva, Cláudio Ricardo da Silva, Rafael Resende Finzi, and Nildo da Silva Dias. Métodos para estimar radiação solar na região noroeste de minas gerais. *Ciência Rural*, 42:276–282, 2012.
- 98 Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.
- 99 Jacyra Soares, Amauri P Oliveira, Marija Zlata Božnar, Primož Mlakar, João F Escobedo, and Antonio J Machado. Modeling hourly diffuse solar-radiation in the city of são paulo using a neural-network technique. *Applied Energy*, 79(2):201–214, 2004. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2003.11.004>. URL <https://www.sciencedirect.com/science/article/pii/S0306261903002381>.
- 100 Adriana Aki Tanaka, Adilson Pacheco de Souza, Antonio Evaldo Klar, Andréa Carvalho da Silva, and Anthony Wellington Almeida Gomes. Evapotranspiração de referência estimada por modelos simplificados para o estado do mato grosso. *Pesquisa agropecuária brasileira*, 51(2):91–104, 2016. ISSN 1678-3921.
- 101 Akin Tascikaraoglu, Borhan M. Sanandaji, Gianfranco Chicco, Valeria Cocina, Filippo Spertino, Ozan Erdinc, Nikolaos G. Paterakis, and João P.S. Catalão. Compressive spatio-temporal forecasting of meteorological quantities and photovoltaic power. *IEEE Transactions on Sustainable Energy*, 7(3):1295–1305, 2016. doi: 10.1109/TSTE.2016.2544929.
- 102 Érico Tadao Teramoto, Darllan Collins da Cunha e Silva, Cícero Manoel Dos Santos, Levi Pompermayer Machado, and João Francisco Escobedo. Estimativa da radiação global diária a partir de variáveis meteorológicas comuns em ambiente subtropical úmido. *Nativa (Sinop)*, 7(6):693, 2019. ISSN 2318-7670.
- 103 Érico Tadao Teramoto, Cícero Manoel dos Santos, João Francisco Escobedo, Alexandre Dal Pai, and Silvia Helena Modenese Gorla da Silva. Comparação de Métodos de Estimativa da Radiação Solar Ultravioleta Horária: Modelos Empíricos, Redes Neurais Artificiais e Máquina de Vetores de Suporte. *Revista Brasileira de Meteorologia*, 35:35 – 43, 03 2020. ISSN 0102-7786. URL http://old.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862020000100035&nrm=iso.
- 104 C. Tiba, R. Reis Reis, J. C. Costa, J. F. Abreu, Melina Amoni, D. P. Guimaraes, and M. A. Porto. On the development of spatial/temporal solar radiation maps: A minas gerais (brazilian) case study. *Journal of Geographic Information System*, 06(03): 258–274, 2014. doi: 10.4236/jgis.2014.63024.
- 105 Lennard Visser, Tarek AlSkaif, and Wilfried van Sark. Operational day-ahead solar power forecasting for aggregated pv systems with a varying spatial distribution. *Renewable Energy*, 183:267–282, 2022. ISSN 0960-1481. doi:

- <https://doi.org/10.1016/j.renene.2021.10.102>. URL
<https://www.sciencedirect.com/science/article/pii/S0960148121015688>.
- 106 Jian Wang, Siyuan Lu, Shui-Hua Wang, and Yu-Dong Zhang. A review on extreme learning machine. *Multimedia Tools and Applications*, May 2021. ISSN 1573-7721. doi: 10.1007/s11042-021-11007-7. URL
<https://doi.org/10.1007/s11042-021-11007-7>.
- 107 Yuhua Wang, Dev Millstein, Andrew D. Mills, Seongeun Jeong, and Amos Ancell. The cost of day-ahead solar forecasting errors in the united states. *Solar Energy*, 231: 846–856, 2022. ISSN 0038-092X. doi: <https://doi.org/10.1016/j.solener.2021.12.012>. URL
<https://www.sciencedirect.com/science/article/pii/S0038092X21010616>.
- 108 Yuan-Kang Wu, Cheng-Liang Huang, Quoc-Thang Phan, and Yuan-Yao Li. Completed review of various solar power forecasting techniques considering different viewpoints. *Energies*, 15(9), 2022. ISSN 1996-1073. doi: 10.3390/en15093320. URL
<https://www.mdpi.com/1996-1073/15/9/3320>.
- 109 Amit Kumar Yadav and S.S. Chandel. Solar radiation prediction using artificial neural network techniques: A review. *Renewable and Sustainable Energy Reviews*, 33: 772–781, 2014. ISSN 1364-0321. doi: <https://doi.org/10.1016/j.rser.2013.08.055>. URL
<https://www.sciencedirect.com/science/article/pii/S1364032113005959>.
- 110 Chen Yang, Anupam A. Thatte, and Le Xie. Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation. *IEEE Transactions on Sustainable Energy*, 6(1):104–112, 2015. doi: 10.1109/TSTE.2014.2359974.
- 111 Dazhi Yang, Chaojun Gu, Zibo Dong, Panida Jirutitijaroen, Nan Chen, and Wilfred M. Walsh. Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. *Renewable Energy*, 60:235–245, 2013. ISSN 0960-1481. doi: <https://doi.org/10.1016/j.renene.2013.05.030>. URL
<https://www.sciencedirect.com/science/article/pii/S0960148113002759>.
- 112 Athanassios Zagouras, Hugo T.C. Pedro, and Carlos F.M. Coimbra. On the role of lagged exogenous variables and spatio-temporal correlations in improving the accuracy of solar forecasting methods. *Renewable Energy*, 78:203–218, 2015. ISSN 0960-1481. doi: <https://doi.org/10.1016/j.renene.2014.12.071>. URL
<https://www.sciencedirect.com/science/article/pii/S0960148115000051>.
- 113 Jianyuan Zhang, Li Zhao, Shuai Deng, Weicong Xu, and Ying Zhang. A critical review of the models used to estimate solar radiation. *Renewable and Sustainable Energy Reviews*, 70:314–329, 2017. ISSN 1364-0321. doi: <https://doi.org/10.1016/j.rser.2016.11.124>. URL
<https://www.sciencedirect.com/science/article/pii/S1364032116308115>.
- 114 Yong Zhou, Yanfeng Liu, Dengjia Wang, Xiaojun Liu, and Yingying Wang. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. *Energy Conversion and Management*, 235:113960, 2021. ISSN 0196-8904. doi: <https://doi.org/10.1016/j.enconman.2021.113960>. URL
<https://www.sciencedirect.com/science/article/pii/S0196890421001369>.

- 115 Cristian Felipe Zuluaga, Alvaro Avila-Diaz, Flavio B. Justino, Fernando Ramos Martins, and Wilmar L. Ceron. The climate change perspective of photovoltaic power potential in brazil. *Renewable Energy*, 193:1019–1031, 2022. ISSN 0960-1481. doi: <https://doi.org/10.1016/j.renene.2022.05.029>. URL <https://www.sciencedirect.com/science/article/pii/S096014812200670X>.

.1 APÊNDICE – Tabelas de dados e resultados

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A422	ABROLHOS	BA	-17 ° 57 ' 47 "	-38 ° 42 ' 12 "	20.93	21/07/2008 30/11/2019	19338
A657	AFONSO CLAUDIO	ES	-20 ° 6 ' 15 "	-41 ° 6 ' 25 "	507.48	28/09/2011 10/12/2021	27980
A045	AGUAS EMENDADAS	DF	-15 ° 35 ' 47 "	-47 ° 37 ' 33 "	1030.36	06/10/2008 31/12/2021	42598
A549	AGUAS VERMELHAS	MG	-15 ° 45 ' 6 "	-41 ° 27 ' 28 "	754.07	09/09/2007 31/12/2021	43876
A534	AIMORES	MG	-19 ° 31 ' 58 "	-41 ° 5 ' 27 "	287.74	21/08/2006 31/12/2021	83311
A617	ALEGRE	ES	-20 ° 45 ' 2 "	-41 ° 29 ' 20 "	138.0	25/10/2006 31/12/2021	46561
A508	ALMENARA	MG	-16 ° 9 ' 60 "	-40 ° 41 ' 16 "	189.11	01/01/2009 29/12/2021	46295
A628	ANGRA DOS REIS	RJ	-22 ° 58 ' 32 "	-44 ° 18 ' 12 "	6.0	25/08/2017 31/12/2021	12756
A566	ARACUAI	MG	-16 ° 50 ' 56 "	-42 ° 2 ' 7 "	308.0	22/05/2017 31/12/2021	12180

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A505	ARAXA	MG	-19 ° 36 ' 21 "	-46 ° 56 ' 59 "	1018.32	01/01/2009 31/12/2019	35049
A736	ARIRANHA	SP	-21 ° 7 ' 59 "	-48 ° 50 ' 26 "	525.44	13/11/2007 31/12/2021	37258
A565	BAMBUI	MG	-20 ° 1 ' 52 "	-46 ° 0 ' 32 "	697.0	24/11/2016 31/12/2021	16155
A502	BARBACENA	MG	-21 ° 13 ' 42 "	-43 ° 46 ' 4 "	1168.76	22/07/2003 09/03/2021	63837
A748	BARRETOS	SP	-20 ° 33 ' 33 "	-48 ° 32 ' 42 "	534.36	22/06/2010 31/12/2021	33592
A447	BELMONTE	BA	-16 ° 5 ' 17 "	-39 ° 12 ' 55 "	89.67	13/07/2009 31/12/2020	25554
A521	BELO HORIZONTE (PAMPULHA)	MG	-19 ° 53 ' 2 "	-43 ° 58 ' 10 "	854.02	01/01/2007 31/12/2021	52251
F501	BELO HORIZONTE - CERCADINHO	MG	-19 ° 58 ' 48 "	-43 ° 57 ' 31 "	1199.55	22/01/2014 31/12/2021	26842
A765	BERTIOGA	SP	-23 ° 50 ' 41 "	-46 ° 8 ' 36 "	5.0	01/02/2017 31/12/2021	10280

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A418	BOM JESUS DA LAPA	BA	-13 ° 15 ' 4 "	-43 ° 24 ' 19 "	447.75	21/05/2007 31/12/2021	39366
A744	BRAGANCA PAULISTA	SP	-22 ° 57 ' 7 "	-46 ° 31 ' 50 "	891.0	20/12/2017 25/12/2021	11925
A001	BRASILIA	DF	-15 ° 47 ' 22 "	-47 ° 55 ' 33 "	1160.96	07/05/2000 31/12/2021	65911
A042	BRAZLANDIA	DF	-15 ° 35 ' 59 "	-48 ° 7 ' 52 "	1143.0	19/07/2017 31/12/2021	14433
A433	BRUMADO	BA	-14 ° 10 ' 55 "	-41 ° 40 ' 20 "	473.37	28/04/2008 13/12/2021	43834
A544	BURITIS	MG	-15 ° 31 ' 27 "	-46 ° 26 ' 8 "	894.0	19/06/2007 01/06/2021	41019
A769	CACHOEIRA PAULISTA	SP	-22 ° 41 ' 20 "	-45 ° 0 ' 20 "	586.0	20/10/2017 30/09/2020	9726
A530	CALDAS	MG	-21 ° 55 ' 5 "	-46 ° 22 ' 59 "	1077.34	28/11/2006 29/10/2021	44321
A604	CAMBUCCI	RJ	-21 ° 35 ' 15 "	-41 ° 57 ' 30 "	46.0	10/12/2002 31/12/2021	39955
A519	CAMPINA VERDE	MG	-19 ° 32 ' 21 "	-49 ° 31 ' 5 "	559.07	15/07/2006 31/12/2021	59168

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A706	CAMPOS DO JORDÃO	SP	-22 ° 45 ' 1 "	-45 ° 36 ' 14 "	1662.95	30/03/2002 25/12/2021	17941
A541	CAPELINHA	MG	-17 ° 42 ' 20 "	-42 ° 23 ' 21 "	932.06	01/09/2007 31/12/2021	42847
A554	CARATINGA	MG	-19 ° 44 ' 9 "	-42 ° 8 ' 14 "	609.25	24/05/2007 31/12/2019	44580
A405	CARAVELAS	BA	-17 ° 44 ' 22 "	-39 ° 15 ' 31 "	6.05	22/12/2002 12/10/2020	39380
A629	CARMO	RJ	-21 ° 56 ' 19 "	-42 ° 36 ' 3 "	293.0	15/10/2018 31/12/2021	8907
A738	CASA BRANCA	SP	-21 ° 46 ' 50 "	-47 ° 4 ' 31 "	734.15	28/06/2007 24/06/2020	39023
A034	CATALÃO	GO	-18 ° 9 ' 17 "	-47 ° 55 ' 39 "	900.72	31/01/2008 31/12/2021	43023
A548	CHAPADA GAUCHA	MG	-15 ° 18 ' 1 "	-45 ° 37 ' 3 "	873.2	23/06/2007 25/09/2021	44371
A520	CONCEICÃO DAS ALAGOAS	MG	-19 ° 59 ' 9 "	-48 ° 9 ' 6 "	572.54	18/07/2006 31/12/2021	46825
A557	CORONEL PACHECO	MG	-21 ° 32 ' 48 "	-43 ° 15 ' 40 "	411.14	19/10/2012 17/09/2021	36048

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A416	CORRENTINA	BA	-13 ° 19 ' 57 "	-44 ° 37 ' 3 "	551.71	09/11/2007 31/12/2021	34864
A036	CRISTALINA	GO	-16 ° 47 ' 6 "	-47 ° 36 ' 47 "	1211.08	16/12/2007 31/12/2021	57868
A056	CRISTALINA (FAZENDA SANTA MONICA)	GO	-16 ° 23 ' 58 "	-47 ° 37 ' 33 "	932.0	29/01/2018 31/12/2021	12422
A538	CURVELO	MG	-18 ° 44 ' 52 "	-44 ° 27 ' 14 "	669.48	19/12/2006 31/12/2021	54945
A537	DIAMANTINA	MG	-18 ° 13 ' 52 "	-43 ° 38 ' 54 "	1359.25	05/06/2007 31/12/2021	45118
A564	DIVINOPOLIS	MG	-20 ° 10 ' 24 "	-44 ° 52 ' 30 "	796.0	30/10/2017 31/12/2021	13090
A536	DORES DO INDAIA	MG	-19 ° 28 ' 55 "	-45 ° 35 ' 38 "	721.09	02/06/2007 31/12/2021	46692
A631	ECOPORANGA	ES	-18 ° 17 ' 29 "	-40 ° 44 ' 11 "	224.0	17/03/2017 31/12/2021	13661
A029	EDEIA	GO	-17 ° 20 ' 13 "	-49 ° 54 ' 53 "	608.0	26/08/2017 31/12/2021	13505

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A543	ESPINOSA	MG	-14 ° 54 ' 44 "	-42 ° 48 ' 31 "	565.36	14/11/2007 19/07/2019	44430
A535	FLORESTAL	MG	-19 ° 53 ' 7 "	-44 ° 25 ' 1 "	753.5	27/06/2008 31/12/2021	58614
A524	FORMIGA	MG	-20 ° 27 ' 18 "	-45 ° 27 ' 14 "	878.14	17/08/2006 31/12/2021	49029
A708	FRANCA	SP	-20 ° 35 ' 4 "	-47 ° 22 ' 57 "	1002.74	12/12/2002 31/12/2021	61856
A532	GOVERNADOR VALADARES	MG	-18 ° 49 ' 49 "	-41 ° 58 ' 37 "	198.24	29/05/2007 31/12/2021	50187
A426	GUANAMBI	BA	-14 ° 12 ' 29 "	-42 ° 44 ' 59 "	552.23	24/04/2008 31/12/2021	37283
A533	GUANHAES	MG	-18 ° 47 ' 13 "	-42 ° 56 ' 35 "	852.68	02/06/2007 31/12/2021	46437
A546	GUARDA-MOR	MG	-17 ° 33 ' 41 "	-47 ° 11 ' 57 "	997.01	12/07/2007 07/04/2021	41796
A555	IBIRITE (ROLA MOCA)	MG	-20 ° 1 ' 53 "	-44 ° 0 ' 40 "	1198.82	06/06/2008 31/12/2021	43596
A455	ITAMARAJU	BA	-17 ° 0 ' 25 "	-39 ° 33 ' 29 "	110.0	04/06/2016 30/10/2020	13401

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A550	ITAOBIM	MG	-16 ° 34 ' 32 "	-41 ° 29 ' 8 "	271.63	05/09/2007 31/12/2021	56462
A446	ITAPETINGA	BA	-15 ° 14 ' 41 "	-40 ° 13 ' 46 "	271.48	25/06/2009 20/12/2021	36190
A739	ITAPIRA	SP	-22 ° 24 ' 54 "	-46 ° 48 ' 19 "	634.89	18/01/2008 16/05/2020	38629
A635	ITATIAIA	RJ	-22 ° 22 ' 26 "	-44 ° 42 ' 11 "	2450.0	01/09/2017 04/11/2018	2174
A512	ITUIUTABA	MG	-18 ° 57 ' 10 "	-49 ° 31 ' 30 "	540.08	12/05/2006 31/12/2021	46132
A035	ITUMBIARA	GO	-18 ° 24 ' 35 "	-49 ° 11 ' 31 "	491.17	01/11/2007 31/12/2021	43513
A753	ITUVERAVA	SP	-20 ° 21 ' 35 "	-47 ° 46 ' 31 "	610.58	18/07/2008 31/03/2021	34248
A733	JALES	SP	-20 ° 9 ' 54 "	-50 ° 35 ' 42 "	460.44	26/08/2007 31/12/2021	44158
A563	JANAUBA	MG	-15 ° 48 ' 10 "	-43 ° 17 ' 49 "	532.0	21/11/2016 31/12/2021	14878
A559	JANUARIA	MG	-15 ° 26 ' 53 "	-44 ° 21 ' 59 "	468.0	27/06/2016 31/12/2021	14419

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A553	JOÃO PINHEIRO	MG	-17 ° 47 ' 4 "	-46 ° 7 ' 10 "	877.05	08/07/2007 31/12/2021	44833
A518	JUIZ DE FORA	MG	-21 ° 46 ' 12 "	-43 ° 21 ' 52 "	936.88	26/05/2007 31/12/2021	50281
A614	LINHARES	ES	-19 ° 21 ' 25 "	-40 ° 4 ' 7 "	40.0	27/10/2006 31/12/2021	59210
A012	LUZIANIA	GO	-16 ° 15 ' 38 "	-47 ° 58 ' 1 "	1000.82	21/10/2006 31/12/2021	47327
A567	MACHADO	MG	-21 ° 40 ' 51 "	-45 ° 56 ' 40 "	969.0	21/07/2017 31/12/2021	14381
A556	MANHUACU	MG	-20 ° 15 ' 48 "	-42 ° 10 ' 58 "	819.47	28/09/2010 31/12/2021	30739
A540	MANTENA	MG	-18 ° 46 ' 50 "	-40 ° 59 ' 11 "	254.91	01/08/2007 31/12/2021	53120
A531	MARIA DA FE	MG	-22 ° 18 ' 52 "	-45 ° 22 ' 23 "	1281.43	01/12/2006 31/12/2021	47938
A632	MARILANDIA	ES	-19 ° 24 ' 26 "	-40 ° 32 ' 23 "	95.0	23/03/2017 31/12/2021	15066
A539	MOCAMBINHO	MG	-15 ° 5 ' 9 "	-44 ° 0 ' 58 "	454.0	11/11/2007 31/12/2021	44451

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A526	MONTALVANIA	MG	-14 ° 24 ' 30 "	-44 ° 24 ' 15 "	519.52	26/06/2007 31/12/2021	44754
A509	MONTE VERDE	MG	-22 ° 51 ' 42 "	-46 ° 2 ' 36 "	1544.89	19/12/2004 31/12/2021	44751
A506	MONTES CLAROS	MG	-16 ° 41 ' 11 "	-43 ° 50 ' 38 "	645.87	19/03/2009 31/12/2021	36801
A003	MORRINHOS	GO	-17 ° 44 ' 42 "	-49 ° 6 ' 6 "	751.09	30/05/2001 31/12/2021	53795
A517	MURIAE	MG	-21 ° 6 ' 18 "	-42 ° 22 ' 33 "	282.79	30/08/2006 31/12/2021	47671
A624	NOVA FRIBURGO	RJ	-22 ° 20 ' 5 "	-42 ° 40 ' 37 "	1070.0	20/09/2010 31/12/2021	30048
A623	NOVA VENECIA	ES	-18 ° 41 ' 43 "	-40 ° 23 ' 26 "	156.02	23/06/2008 31/12/2021	47623
A570	OLIVEIRA	MG	-20 ° 42 ' 54 "	-44 ° 51 ' 52 "	1025.0	28/11/2017 31/12/2021	7922
A513	OURO BRANCO	MG	-20 ° 33 ' 24 "	-43 ° 45 ' 22 "	1048.2	28/07/2006 31/12/2021	51928
A571	PARACATU	MG	-17 ° 14 ' 40 "	-46 ° 52 ' 54 "	705.0	16/03/2018 31/12/2021	8801

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A710	PARANAIBA	MS	-19 ° 41 ' 44 ''	-51 ° 10 ' 54 ''	408.06	19/12/2006 01/11/2021	28545
A047	PARANOIA (COOPA-DF)	DF	-16 ° 0 ' 44 ''	-47 ° 33 ' 27 ''	1043.0	09/02/2017 31/12/2021	14937
A619	PARATI	RJ	-23 ° 13 ' 25 ''	-44 ° 43 ' 37 ''	3.0	06/12/2006 31/12/2021	48205
A529	PASSA QUATRO	MG	-22 ° 23 ' 45 ''	-44 ° 57 ' 43 ''	1017.1	04/06/2007 31/12/2021	43780
A516	PASSOS	MG	-20 ° 44 ' 43 ''	-46 ° 38 ' 2 ''	781.7	18/07/2006 31/12/2021	41926
A562	PATOS DE MINAS	MG	-18 ° 31 ' 14 ''	-46 ° 26 ' 26 ''	950.0	12/05/2017 07/09/2021	12028
A523	PATROCINIO	MG	-18 ° 59 ' 48 ''	-46 ° 59 ' 9 ''	978.11	22/08/2006 31/12/2021	47094
A610	PICO DO COUTO	RJ	-22 ° 27 ' 53 ''	-43 ° 17 ' 29 ''	1777.0	28/10/2006 30/12/2021	38932
A545	PIRAPORA	MG	-17 ° 15 ' 29 ''	-44 ° 50 ' 8 ''	505.32	04/07/2007 31/12/2021	45298
A033	PIRES DO RIO	GO	-17 ° 18 ' 15 ''	-48 ° 17 ' 3 ''	757.27	12/10/2007 31/12/2021	44212

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A427	PORTO SEGURO	BA	-16 ° 23 ' 20 "	-39 ° 10 ' 57 "	86.36	11/07/2008 04/10/2020	40273
A017	POSSE	GO	-14 ° 5 ' 21 "	-46 ° 21 ' 59 "	830.0	20/04/2007 09/12/2021	42249
A747	PRADOPOLIS	SP	-21 ° 20 ' 18 "	-48 ° 6 ' 50 "	540.41	22/04/2008 31/12/2021	52639
A622	PRES. KENNEDY	ES	-21 ° 6 ' 3 "	-41 ° 2 ' 22 "	69.12	25/06/2008 31/12/2021	48143
A609	RESENDE	RJ	-22 ° 27 ' 5 "	-44 ° 26 ' 42 "	438.83	29/09/2006 31/12/2021	46600
A626	RIO CLARO	RJ	-22 ° 39 ' 13 "	-44 ° 2 ' 27 "	516.0	07/06/2016 31/12/2021	15235
A551	RIO PARDO DE MINAS	MG	-15 ° 43 ' 23 "	-42 ° 26 ' 9 "	850.06	19/11/2007 31/12/2021	56410
A525	SACRAMENTO	MG	-19 ° 52 ' 31 "	-47 ° 26 ' 3 "	913.12	19/08/2006 31/12/2021	47281
A552	SALINAS	MG	-16 ° 9 ' 37 "	-42 ° 18 ' 37 "	487.14	14/09/2007 31/12/2021	54553
A630	SANTA MARIA MADALENA	RJ	-21 ° 57 ' 2 "	-42 ° 0 ' 37 "	517.0	16/10/2018 31/12/2021	9646

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A613	SANTA TERESA	ES	-19 ° 59 ' 18 "	-40 ° 34 ' 46 "	975.92	09/08/2007 21/12/2020	41432
A711	SÃO CARLOS	SP	-21 ° 58 ' 49 "	-47 ° 53 ' 2 "	859.29	05/09/2006 31/12/2021	45581
A514	SÃO JOÃO DEL REI	MG	-21 ° 6 ' 23 "	-44 ° 15 ' 3 "	929.88	10/06/2006 31/12/2021	49930
A740	SÃO LUIS DO PARAÍTINGA	SP	-23 ° 13 ' 42 "	-45 ° 25 ' 1 "	862.3	05/11/2007 22/04/2021	38679
A616	SÃO MATEUS	ES	-18 ° 40 ' 34 "	-39 ° 51 ' 51 "	28.66	25/10/2006 31/12/2021	64166
A701	SÃO PAULO - MIRANTE	SP	-23 ° 29 ' 47 "	-46 ° 37 ' 12 "	785.64	25/07/2006 31/12/2021	50194
A547	SÃO ROMÃO	MG	-16 ° 21 ' 46 "	-45 ° 7 ' 26 "	490.29	30/06/2007 31/12/2021	44836
A561	SÃO SEBASTIÃO DO PARAISO	MG	-20 ° 54 ' 36 "	-47 ° 6 ' 51 "	845.0	18/08/2015 10/01/2021	25820
A522	SERRA DOS AIMORES	MG	-17 ° 47 ' 56 "	-40 ° 15 ' 0 "	211.97	21/08/2006 30/10/2021	50895
A569	SETE LAGOAS	MG	-19 ° 27 ' 19 "	-44 ° 10 ' 24 "	719.0	10/06/2016 31/12/2021	12586

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A728	TAUBATE	SP	-23 ° 2 ' 30 "	-45 ° 31 ' 15 "	582.26	20/12/2006 26/02/2021	36043
A527	TEOFILO OTONI	MG	-17 ° 53 ' 34 "	-41 ° 30 ' 56 "	467.17	25/08/2006 31/12/2021	44265
A618	TERESOPOLIS	RJ	-22 ° 26 ' 55 "	-42 ° 59 ' 13 "	981.0	01/11/2006 31/12/2021	52898
A511	TIMOTEO	MG	-19 ° 34 ' 26 "	-42 ° 37 ' 21 "	493.42	24/02/2006 31/12/2021	46956
A528	TRES MARIAS	MG	-18 ° 12 ' 3 "	-45 ° 27 ' 35 "	931.01	25/08/2006 31/12/2021	54145
A625	TRES RIOS	RJ	-22 ° 5 ' 54 "	-43 ° 12 ' 30 "	295.0	10/06/2016 31/12/2021	14120
A568	UBERABA	MG	-19 ° 42 ' 36 "	-47 ° 57 ' 43 "	778.0	19/05/2017 31/12/2021	12935
A507	UBERLANDIA	MG	-18 ° 55 ' 1 "	-48 ° 15 ' 20 "	874.77	21/03/2003 31/12/2021	66161
A542	UNAI	MG	-16 ° 33 ' 15 "	-46 ° 52 ' 55 "	640.85	16/06/2007 31/12/2021	46177
A611	VALENCA	RJ	-22 ° 21 ' 29 "	-43 ° 41 ' 44 "	370.0	27/09/2006 31/12/2021	47454

Continua na próxima página

Tabela Suplementar S1 – Estações meteorológicas automáticas

Código da Estação	Nome da Cidade	Unidade Federativa	Latitude (S)	Longitude (W)	Altitude (m)	Intervalo dos Dados	Entradas Totais
A515	VARGINHA	MG	-21 ° 33 ' 59 "	-45 ° 24 ' 15 "	949.78	13/07/2006 31/12/2021	48052
A633	VENDA NOVA DO IMIGRANTE	ES	-20 ° 23 ' 8 "	-41 ° 11 ' 24 "	722.0	10/02/2017 17/01/2021	6627
A510	VICOSA	MG	-20 ° 45 ' 45 "	-42 ° 51 ' 50 "	697.64	15/09/2005 31/12/2021	61205
A634	VILA VELHA	ES	-20 ° 28 ' 1 "	-40 ° 24 ' 14 "	25.0	16/02/2017 31/12/2021	13593
A612	VITORIA	ES	-20 ° 16 ' 15 "	-40 ° 18 ' 21 "	9.0	31/10/2006 31/12/2021	53970
A414	VITORIA DA CONQUISTA	BA	-14 ° 53 ' 11 "	-40 ° 48 ' 5 "	879.38	01/06/2007 31/12/2021	35107
A729	VOTUPORANGA	SP	-20 ° 24 ' 12 "	-49 ° 57 ' 58 "	510.37	08/12/2006 28/02/2021	40952

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
AGUAS VERMELHAS	1	0.814	0.947	2.71	2.14	0.425
	2	0.798	0.946	2.85	2.25	0.514
	3	0.665	0.924	3.66	2.90	-0.378
	4	0.793	0.945	2.88	2.26	0.755
	5	0.766	0.941	3.06	2.39	0.930
AIMORES	1	0.635	0.898	7.99	5.70	2.47
	2	0.628	0.891	8.14	5.84	2.58
	3	0.387	0.796	10.19	7.99	-1.45
	4	0.630	0.892	8.11	5.79	2.64
	5	0.293	0.744	10.89	8.67	-5.27
ALMENARA	1	0.753	0.927	3.28	2.58	-0.125
	2	0.741	0.921	3.37	2.64	-0.148
	3	0.675	0.899	3.77	2.94	-0.697
	4	0.709	0.904	3.56	2.79	-0.098
	5	0.681	0.896	3.73	2.92	-0.475
ARACUAI	1	0.547	0.834	4.52	3.16	1.54
	2	0.138	0.778	5.85	4.15	0.342
	3	0.121	0.776	5.90	4.16	0.106
	4	0.191	0.785	5.60	3.93	0.496
	5	0.195	0.785	5.60	3.92	0.395
ARAXA	1	0.537	0.824	4.22	2.34	0.443
	2	0.484	0.780	4.58	2.42	0.457
	3	0.484	0.781	4.58	2.42	0.048
	4	0.488	0.783	4.57	2.40	0.408
	5	0.486	0.782	4.58	2.41	-0.037
BAMBUI	1	0.614	0.893	3.26	2.33	0.114
	2	0.703	0.913	2.99	2.31	0.277
	3	0.690	0.909	3.06	2.39	-0.234
	4	0.724	0.919	2.89	2.28	-0.522
	5	0.711	0.915	2.96	2.33	-0.539
BARBACENA	1	0.445	0.867	4.26	2.68	0.267
	2	0.421	0.860	4.35	2.80	-0.028
	3	0.422	0.863	4.36	3.03	-1.36
	4	0.493	0.869	4.07	2.77	-0.368
	5	0.452	0.861	4.24	2.91	-0.911
BELO HORIZONTE PAMPULHA	1	0.806	0.942	2.30	1.81	0.513
	2	0.752	0.923	2.60	2.05	1.08
	3	-17618.17	0.298	560.78	560.55	559.60
	4	0.728	0.915	2.72	2.09	1.21
	5	0.596	0.872	3.32	2.53	1.84
BELO HORIZONTE	1	0.731	0.921	2.80	2.22	-0.146
	2	0.195	0.833	4.69	3.23	-0.822
	3	0.042	0.817	5.13	3.69	0.372

Continua na próxima página

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
CERCADINHO	4	0.566	0.893	3.47	2.59	-0.321
	5	0.460	0.877	3.86	2.87	-0.342
BURITIS	1	0.713	0.910	3.05	2.29	0.948
	2	0.739	0.924	2.96	2.23	0.693
	3	0.732	0.920	2.99	2.30	0.213
	4	0.746	0.923	2.92	2.18	0.509
	5	0.742	0.921	2.94	2.21	0.289
CALDAS	1	0.601	0.894	3.36	2.01	0.645
	2	0.628	0.902	3.21	2.12	0.922
	3	0.434	0.871	3.84	2.82	-0.764
	4	0.696	0.916	2.91	2.14	0.682
	5	0.640	0.900	3.17	2.41	0.160
CAMPINA VERDE	1	0.525	0.850	3.38	2.47	1.24
	2	0.608	0.870	3.04	2.22	0.967
	3	0.651	0.885	2.87	2.16	0.202
	4	0.585	0.866	3.13	2.28	1.03
	5	0.610	0.876	3.02	2.25	0.434
CAPELINHA	1	0.401	0.852	4.58	2.87	0.269
	2	0.235	0.826	5.19	3.08	0.596
	3	-0.950	0.702	8.20	5.17	-0.631
	4	0.142	0.817	5.46	3.27	0.466
	5	-0.863	0.707	7.99	4.83	4.21
CARATINGA	1	0.483	0.820	5.06	3.35	1.42
	2	0.271	0.742	6.16	4.16	2.83
	3	0.019	0.707	7.13	5.16	4.57
	4	0.277	0.730	6.13	4.21	2.66
	5	0.206	0.726	6.43	4.47	3.26
CHAPADA GAUCHA	1	0.711	0.905	2.85	2.19	-0.141
	2	0.693	0.902	3.01	2.32	-0.140
	3	0.687	0.902	3.04	2.31	0.018
	4	0.697	0.904	2.97	2.30	-0.348
	5	0.697	0.905	2.98	2.32	-0.352
CONCEIÇÃO DAS ALAGOAS	1	0.664	0.895	2.85	2.18	1.10
	2	0.660	0.895	2.88	2.18	1.06
	3	-0.694	0.835	4.49	3.90	0.426
	4	0.640	0.890	2.96	2.22	1.35
	5	0.459	0.851	3.61	2.83	2.38
CORONEL PACHECO	1	0.777	0.933	2.68	2.12	-0.165
	2	0.678	0.897	3.44	2.61	1.33
	3	0.382	0.840	4.65	3.78	3.24
	4	0.654	0.892	3.56	2.69	0.904
	5	0.269	0.808	5.17	4.30	3.67
CURVELO	1	0.752	0.928	2.97	2.18	0.712
	2	0.482	0.853	4.02	2.73	1.07

Continua na próxima página

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
	3	-4.45	0.579	12.09	10.64	4.42
	4	0.522	0.851	3.93	2.71	1.15
	5	0.546	0.842	4.01	2.88	0.821
DIAMANTINA	1	0.692	0.905	3.33	2.30	0.411
	2	0.705	0.913	3.25	2.34	0.430
	3	0.199	0.855	4.69	3.99	-2.65
	4	0.695	0.910	3.30	2.45	-0.053
	5	0.246	0.855	4.53	3.86	-2.56
DIVINOPOLIS	1	0.804	0.946	2.30	1.82	0.279
	2	0.690	0.910	2.91	2.24	-0.095
	3	0.641	0.898	3.13	2.43	0.861
	4	0.776	0.934	2.50	1.91	-0.359
	5	0.776	0.934	2.50	1.91	-0.359
DORES DO INDAIA	1	0.817	0.946	2.26	1.69	0.414
	2	0.813	0.945	2.28	1.70	0.429
	3	-3.82	0.685	8.60	7.57	4.30
	4	0.797	0.939	2.34	1.72	0.378
	5	0.745	0.923	2.62	1.94	0.573
ESPINOSA	1	0.734	0.917	2.76	2.04	0.433
	2	0.622	0.862	3.41	2.53	0.052
	3	0.316	0.763	4.59	3.54	-1.90
	4	0.618	0.859	3.43	2.57	-0.218
	5	0.484	0.810	3.98	3.08	-1.06
FLORESTAL	1	0.668	0.917	2.82	2.12	1.03
	2	0.547	0.890	3.29	2.49	1.88
	3	0.331	0.857	3.97	2.97	1.44
	4	0.539	0.887	3.32	2.55	1.96
	5	0.392	0.865	3.74	2.88	1.76
FORMIGA	1	0.799	0.940	2.36	1.83	0.086
	2	0.799	0.940	2.36	1.83	0.088
	3	0.683	0.911	2.81	2.23	1.13
	4	0.795	0.938	2.38	1.84	0.175
	5	-11.80	0.848	7.08	6.51	-3.43
GOVERNADOR VALADARES	1	0.752	0.925	3.10	2.46	-0.581
	2	0.719	0.916	3.30	2.63	-1.40
	3	0.738	0.925	3.19	2.43	0.249
	4	0.718	0.909	3.28	2.54	-0.553
	5	0.744	0.921	3.15	2.43	0.778
GUANHAES	1	0.822	0.950	2.52	1.90	0.619
	2	0.773	0.934	2.83	2.07	0.705
	3	0.695	0.909	3.29	2.41	0.806
	4	0.786	0.935	2.74	2.06	0.552
	5	0.780	0.930	2.80	2.15	0.344
	1	0.746	0.922	2.71	2.04	0.154

Continua na próxima página

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
GUARDA-MOR	2	0.748	0.920	2.79	2.14	-0.217
	3	-93.57	0.865	12.97	12.29	9.71
	4	0.732	0.917	2.87	2.22	-0.228
	5	0.606	0.884	3.45	2.72	-1.30
	1	0.767	0.929	2.59	2.05	-0.122
IBIRITE (ROLA MOCA)	2	0.751	0.924	2.74	2.16	-0.274
	3	0.450	0.853	4.05	3.36	-2.61
	4	0.718	0.910	2.90	2.33	-0.494
	5	0.644	0.883	3.26	2.69	-1.23
	1	0.785	0.936	2.91	2.31	-0.029
ITAOBIM	2	0.746	0.916	3.16	2.51	0.035
	3	0.427	0.786	4.64	3.92	-2.30
	4	0.742	0.916	3.19	2.58	-0.461
	5	0.697	0.901	3.45	2.88	-1.15
	1	0.309	0.818	3.95	3.07	2.21
ITUIUTABA	2	0.641	0.886	2.90	2.22	0.810
	3	0.650	0.891	2.86	2.19	0.675
	4	0.614	0.872	3.02	2.28	1.02
	5	0.551	0.856	3.24	2.49	1.26
	1	0.455	0.860	3.68	2.87	0.592
JANAUBA	2	0.667	0.894	3.12	2.40	-0.160
	3	0.630	0.881	3.28	2.59	-0.622
	4	0.623	0.886	3.30	2.60	-0.890
	5	0.602	0.881	3.39	2.67	-0.951
	1	0.626	0.889	3.05	2.22	-0.119
JANUARIA	2	0.653	0.899	3.09	2.28	-0.276
	3	0.649	0.897	3.11	2.31	-0.296
	4	0.629	0.893	3.20	2.34	-0.052
	5	0.624	0.892	3.22	2.36	-0.028
	1	0.718	0.913	2.65	1.94	0.665
JOÃO PINHEIRO	2	0.708	0.911	2.67	1.95	0.623
	3	0.653	0.890	2.90	2.21	0.140
	4	0.674	0.896	2.82	2.05	0.656
	5	0.574	0.861	3.21	2.38	0.775
	1	0.843	0.955	2.41	1.89	-0.051
JUIZ DE FORA	2	0.826	0.949	2.56	2.01	-0.124
	3	-0.662	0.715	7.41	6.71	-6.25
	4	0.808	0.942	2.71	2.15	-0.484
	5	0.619	0.893	3.76	3.09	-2.25
	1	0.755	0.928	2.92	2.16	0.184
MACHADO	2	0.746	0.920	2.88	2.33	-0.840
	3	0.746	0.920	2.88	2.33	-0.840
	4	0.711	0.909	3.07	2.55	-1.31

Continua na próxima página

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
MANHUACU	5	0.711	0.909	3.07	2.55	-1.31
	1	0.743	0.925	3.57	2.85	-1.78
	2	0.648	0.888	4.25	3.36	-2.33
	3	0.162	0.773	6.55	5.57	-5.28
	4	0.562	0.862	4.76	3.83	-3.11
MANTENA	5	0.445	0.829	5.35	4.40	-3.90
	1	0.836	0.949	2.52	1.99	-0.261
	2	0.798	0.933	2.77	2.22	-0.599
	3	0.634	0.870	3.70	2.94	-0.182
	4	0.776	0.926	2.89	2.29	-0.530
MARIA DA FE	5	0.517	0.841	4.22	3.33	0.358
	1	0.286	0.754	5.98	3.25	1.51
	2	0.342	0.774	5.51	3.06	1.33
	3	0.220	0.729	5.89	3.97	0.034
	4	0.342	0.774	5.51	3.06	1.33
MOCAMBINHO	5	0.220	0.729	5.89	3.97	0.034
	1	0.728	0.914	2.67	2.00	0.131
	2	0.724	0.911	2.69	2.01	0.230
	3	0.683	0.899	2.88	2.17	0.243
	4	0.735	0.916	2.64	2.01	0.049
MONTALVANIA	5	0.714	0.910	2.74	2.15	-0.346
	1	0.438	0.793	4.86	3.75	2.11
	2	0.444	0.800	4.80	3.69	1.94
	3	0.448	0.795	4.78	3.78	1.38
	4	0.418	0.779	4.91	3.78	2.14
MONTE VERDE	5	0.357	0.763	5.16	3.95	2.52
	1	0.620	0.895	3.69	2.80	0.937
	2	0.668	0.904	3.48	2.65	0.768
	3	-217.65	0.792	29.42	28.73	15.93
	4	0.578	0.874	3.93	3.02	1.22
MONTES CLAROS	5	-0.729	0.773	6.35	5.32	-3.29
	1	0.719	0.908	2.83	2.24	-0.370
	2	0.673	0.892	3.04	2.35	0.171
	3	-5.04	0.818	6.23	5.44	-2.67
	4	0.645	0.882	3.17	2.44	0.314
MURIAE	5	0.570	0.864	3.48	2.62	0.793
	1	0.797	0.944	2.73	2.13	1.34
	2	0.778	0.939	2.91	2.25	1.26
	3	0.673	0.917	3.51	2.75	1.83
	4	0.789	0.943	2.84	2.20	1.23
OLIVEIRA	5	0.713	0.926	3.29	2.55	1.55
	1	0.704	0.922	2.66	2.01	0.403
	2	-92.51	0.790	20.85	20.35	-2.47
	3	-92.51	0.790	20.85	20.35	-2.47

Continua na próxima página

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
	4	0.730	0.923	2.80	2.25	-1.31
	5	0.730	0.923	2.80	2.25	-1.31
OURO BRANCO	1	0.647	0.886	3.00	2.35	-0.157
	2	0.622	0.882	3.09	2.42	-0.212
	3	-1738.78	0.552	116.41	115.88	114.68
	4	0.614	0.889	3.18	2.51	0.432
	5	0.502	0.873	3.61	2.88	1.54
PARACATU	1	0.587	0.875	3.59	2.80	0.496
	2	0.561	0.867	3.51	2.57	1.46
	3	0.561	0.867	3.51	2.57	1.46
	4	0.577	0.865	3.45	2.50	1.65
	5	0.577	0.865	3.45	2.50	1.65
PASSA QUATRO	1	0.747	0.926	3.04	2.33	0.029
	2	0.689	0.906	3.46	2.50	0.188
	3	-12.48	0.722	10.09	9.02	-1.81
	4	0.683	0.903	3.47	2.48	0.686
	5	0.552	0.871	4.12	3.17	2.00
PASSOS	1	0.749	0.922	2.81	2.09	0.464
	2	0.725	0.913	2.91	2.17	0.535
	3	0.434	0.883	3.56	2.76	0.509
	4	0.724	0.909	2.92	2.20	0.558
	5	0.649	0.889	3.27	2.55	0.719
PATOS DE MINAS	1	0.855	0.958	2.61	2.06	0.091
	2	0.823	0.948	2.75	2.06	0.731
	3	0.829	0.951	2.70	2.03	0.683
	4	0.818	0.947	2.79	2.09	0.938
	5	0.819	0.948	2.78	2.09	0.915
PATROCINIO	1	0.735	0.921	2.45	1.89	0.377
	2	0.708	0.916	2.57	1.99	0.567
	3	0.159	0.823	4.33	3.42	2.35
	4	0.687	0.912	2.66	2.07	0.670
	5	0.250	0.836	4.07	3.28	2.37
PIRAPORA	1	0.683	0.893	3.19	2.37	0.196
	2	0.683	0.889	3.18	2.39	0.082
	3	0.625	0.868	3.44	2.68	-0.251
	4	0.682	0.887	3.13	2.39	0.075
	5	0.638	0.874	3.30	2.56	0.491
RIO PARDO DE MINAS	1	0.823	0.951	2.50	1.97	-0.326
	2	0.814	0.950	2.56	2.00	0.038
	3	0.548	0.889	3.93	3.24	-1.93
	4	0.804	0.945	2.62	2.09	-0.516
	5	0.760	0.933	2.90	2.29	-0.186
SACRAMENTO	1	0.674	0.897	3.15	2.30	1.00
	2	0.610	0.879	3.46	2.50	1.37

Continua na próxima página

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
	3	-1.76	0.748	6.64	5.12	3.45
	4	0.602	0.862	3.51	2.49	1.12
	5	0.471	0.826	4.03	2.84	1.64
SALINAS	1	0.749	0.928	3.12	2.41	0.006
	2	0.715	0.912	3.33	2.59	-0.212
	3	0.685	0.901	3.51	2.72	0.261
	4	0.737	0.918	3.21	2.50	0.485
	5	0.742	0.918	3.18	2.51	0.661
SÃO JOÃO DEL REI	1	0.792	0.940	2.48	1.91	0.197
	2	0.758	0.932	2.68	2.01	0.248
	3	-197.87	0.692	29.92	29.27	-0.190
	4	0.715	0.916	2.91	2.19	0.632
	5	0.364	0.839	4.26	3.54	2.91
SÃO ROMÃO	1	0.602	0.868	3.31	2.34	1.05
	2	0.614	0.870	3.33	2.42	1.04
	3	0.631	0.875	3.26	2.41	0.785
	4	0.624	0.871	3.29	2.44	0.927
	5	0.618	0.869	3.32	2.51	0.730
SÃO SEBASTIÃO DO PARAISO	1	0.683	0.900	2.90	2.17	-0.359
	2	0.730	0.917	2.77	2.11	0.139
	3	0.725	0.916	2.79	2.12	0.210
	4	0.695	0.904	2.94	2.32	-0.779
	5	0.695	0.904	2.94	2.33	-0.798
SERRA DOS AIMORES	1	0.833	0.952	2.52	1.97	0.373
	2	0.808	0.944	2.70	2.07	0.273
	3	0.646	0.898	3.62	2.85	-0.553
	4	0.783	0.934	2.86	2.19	0.600
	5	-0.609	0.705	7.67	7.17	-6.94
SETE LAGOAS	1	0.432	0.851	3.58	2.87	2.02
	2	0.515	0.865	3.49	2.76	1.72
	3	0.316	0.825	4.07	3.34	2.67
	4	0.622	0.882	3.09	2.43	0.731
	5	0.536	0.863	3.40	2.72	0.068
TEOFILO OTONI	1	0.232	0.809	5.63	3.12	1.98
	2	0.572	0.877	4.26	2.65	1.35
	3	-31.42	0.771	13.96	12.86	-8.95
	4	0.799	0.942	2.93	2.23	1.03
	5	0.809	0.944	2.87	2.26	0.261
TIMOTEO	1	0.817	0.947	2.37	1.88	0.314
	2	0.803	0.945	2.45	1.94	0.130
	3	0.711	0.924	2.96	2.36	-0.606
	4	0.775	0.936	2.62	2.08	0.363
	5	0.645	0.903	3.29	2.63	1.36
	1	0.740	0.920	2.67	1.99	0.562

Continua na próxima página

Tabela Suplementar S2 – Resultados

Estação meteorológica	Cenário	R ²	WI	RMSE	MAE	MBE
TRES	2	0.738	0.922	2.67	2.00	0.695
MARIAS	3	0.681	0.909	2.93	2.25	0.814
	4	0.732	0.919	2.70	2.05	0.065
	5	0.708	0.915	2.82	2.22	-0.513
UBERABA	1	0.119	0.647	7.72	4.74	3.05
	2	0.411	0.754	5.19	3.47	-0.687
	3	0.411	0.754	5.19	3.47	-0.687
	4	0.402	0.753	5.22	3.31	-0.071
	5	0.402	0.753	5.22	3.31	-0.071
UBERLANDIA	1	0.624	0.890	2.88	2.17	1.44
	2	0.586	0.868	2.99	2.17	1.38
	3	-3.44	0.822	5.14	4.57	-2.71
	4	0.560	0.863	3.08	2.21	1.45
	5	0.617	0.875	2.87	2.15	0.498
UNAI	1	0.674	0.899	3.08	2.15	0.091
	2	0.676	0.899	3.06	2.15	0.069
	3	0.556	0.866	3.48	2.76	-1.39
	4	0.658	0.890	3.15	2.24	0.063
	5	0.491	0.826	3.81	2.90	-0.290
VARGINHA	1	0.761	0.928	2.50	1.90	0.866
	2	0.750	0.925	2.57	1.94	0.936
	3	-128.98	0.723	18.12	17.19	16.13
	4	0.694	0.909	2.85	2.24	0.254
	5	0.253	0.808	4.36	3.70	-2.04
VICOSA	1	0.846	0.956	2.17	1.69	-0.101
	2	0.840	0.956	2.21	1.70	0.202
	3	0.762	0.939	2.65	2.11	-0.765
	4	0.832	0.951	2.26	1.76	-0.096
	5	0.810	0.944	2.40	1.87	-0.427

Tabela Suplementar S3 – Comparação com literatura

Estação meteorológica	Artigo	Modelo de Aprendizado	R ²
AGUAS	Este trabalho	ELM	0.814
	(33)	Empírico	0.5657
	VERMELHAS	(25)	Empírico
(13)		MARS	0.702
AIMORES	Este trabalho	ELM	0.635
	Cunha et al. (33)	Empírico	0.4940
	Silva et al. (96)*	Empírico	0.7336
	(13)	MARS	0.637

Continua na próxima página

Tabela Suplementar S3 – Comparação com literatura

Estação meteorológica	Artigo	Modelo de Aprendizado	R ²
ALMENARA	Este trabalho	ELM	0.753
	Cunha et al. (33)	Empírico	0.5054
	Paula et al. (86) (13)	Empírico MARS	0.7272 0.777
ARACUAI	Este trabalho	ELM	0.547
ARAXA	Este trabalho	ELM	0.537
	Cunha et al. (33)	Empírico	0.4597
	Ramos et al. (91) (13)	Empírico MARS	0.55 0.694
BAMBUI	Este trabalho	ELM	0.614
BARBACENA	Este trabalho	ELM	0.445
	Cunha et al. (33)	Empírico	0.4618
	Barbosa et al. (8)* (13)	Empírico MARS	0.72 0.640
BELO HORIZONTE	Este trabalho	ELM	0.806
	Cunha et al. (33)	Empírico	0.4610
	Silva et al. (96)*	Empírico	0.7336
PAMPULHA	Ramos et al. (91) (13)	Empírico MARS	0.51 0.791
	BELO HORIZONTE CERCADINHO	Este trabalho	ELM
BURITIS	Este trabalho	ELM	0.713
	Cunha et al. (33)	Empírico	0.3967
	Silva et al. (97)*	Empírico	0.62
	Finzi et al. (51)* (13)	Empírico MARS	0.62 0.577
CALDAS	Este trabalho	ELM	0.601
	Cunha et al. (33)	Empírico	0.5462
	Barbosa et al. (8)* (13)	Empírico MARS	0.72 0.825
CAMPINA VERDE	Este trabalho	ELM	0.525
	Cunha et al. (33) (13)	Empírico MARS	0.3684 0.606
CAPELINHA	Este trabalho	ELM	0.401
	Cunha et al. (33)	Empírico	0.5958
	Paula et al. (86) (13)	Empírico MARS	0.7272 0.754
CARATINGA	Este trabalho	ELM	0.483
	Cunha et al. (33)	Empírico	0.4693
	Silva et al. (96)* (13)	Empírico MARS	0.7336 0.726
CHAPADA	Este trabalho	ELM	0.711
	Cunha et al. (33)	Empírico	0.5023

Continua na próxima página

Tabela Suplementar S3 – Comparação com literatura

Estação meteorológica	Artigo	Modelo de Aprendizado	R ²
GAUCHA	Carvalho et al. (25)*	Empírico	0.711
	(13)	MARS	0.658
CONCEIÇÃO	Este trabalho	ELM	0.664
DAS	Cunha et al. (33)	Empírico	0.3948
ALAGOAS	(13)	MARS	0.722
CORONEL	Este trabalho	ELM	0.777
PACHECO	Este trabalho	ELM	0.752
CURVELO	Cunha et al. (33)	Empírico	0.4592
	(13)	MARS	0.758
DIAMANTINA	Este trabalho	ELM	0.692
	Cunha et al. (33)	Empírico	0.5955
	Paula et al. (86)	Empírico	0.7272
	(13)	MARS	0.747
DIVINOPOLIS	Este trabalho	ELM	0.804
DORES	Este trabalho	ELM	0.817
DO	Cunha et al. (33)	Empírico	0.4847
INDAIA	(13)	MARS	0.789
ESPINOSA	Este trabalho	ELM	0.734
	Cunha et al. (33)	Empírico	0.3967
	Carvalho et al. (25)*	Empírico	0.711
	(13)	MARS	0.577
FLORESTAL	Este trabalho	ELM	0.668
	Cunha et al. (33)	Empírico	0.4817
	Silva et al. (96)*	Empírico	0.7336
	(13)	MARS	0.741
FORMIGA	Este trabalho	ELM	0.799
	Cunha et al. (33)	Empírico	0.4787
	Barbosa et al. (8)*	Empírico	0.72
	(13)	MARS	0.809
GOVERNADOR	Este trabalho	ELM	0.752
	Cunha et al. (33)	Empírico	0.4535
VALADARES	Silva et al. (96)*	Empírico	0.7336
	(13)	ANN	0.697
GUANHAES	Este trabalho	ELM	0.822
	Cunha et al. (33)	Empírico	0.5899
	(13)	MARS	0.867
GUARDA-MOR	Este trabalho	ELM	0.746
	Cunha et al. (33)	Empírico	0.3988
	Silva et al. (97)*	Empírico	0.62
	Finzi et al. (51)*	Empírico	0.62
	(13)	MARS	0.683
IBIRITE	Este trabalho	ELM	0.767
	Cunha et al. (33)	Empírico	0.4998

Continua na próxima página

Tabela Suplementar S3 – Comparação com literatura

Estação meteorológica	Artigo	Modelo de Aprendizado	R ²
(ROLA MOCA)	Silva et al. (96)*	Empírico	0.7336
	(13)	MARS	0.660
ITAOBIM	Este trabalho	ELM	0.785
	Cunha et al. (33)	Empírico	0.4884
	Paula et al. (86) (13)	Empírico ANN	0.7272 0.683
ITUIUTABA	Este trabalho	ELM	0.309
	Cunha et al. (33) (13)	Empírico MARS	0.3847 0.589
JANAUBA	Este trabalho	ELM	0.455
JANUARIA	Este trabalho	ELM	0.626
JOÃO PINHEIRO	Este trabalho	ELM	0.718
	Cunha et al. (33)	Empírico	0.5464
	Silva et al. (97)* Finzi et al. (51)* (13)	Empírico Empírico MARS	0.62 0.62 0.710
JUIZ DE FORA	Este trabalho	ELM	0.843
	Cunha et al. (33)	Empírico	0.5443
	Silva et al. (96)* (13)	Empírico MARS	0.7336 0.824
MACHADO	Este trabalho	ELM	0.755
MANHUACU	Este trabalho	ELM	0.743
MANTENA	Este trabalho	ELM	0.836
	Cunha et al. (33)	Empírico	0.4831
	Silva et al. (96)* (13)	Empírico MARS	0.7336 0.738
MARIA DA FE	Este trabalho	ELM	0.286
	Cunha et al. (33)	Empírico	0.5251
	Barbosa et al. (8)* (13)	Empírico MARS	0.72 0.828
MOCAMBINHO	Este trabalho	ELM	0.728
	Cunha et al. (33)	Empírico	0.4459
	Carvalho et al. (25)* (13)	Empírico MARS	0.711 0.666
MONTALVANIA	Este trabalho	ELM	0.438
	Cunha et al. (33)	Empírico	0.3162
	Carvalho et al. (25)* (13)	Empírico MARS	0.711 0.544
MONTE VERDE	Este trabalho	ELM	0.620
	Cunha et al. (33) (13)	Empírico MARS	0.4635 0.798
MONTES	Este trabalho	ELM	0.719
	Cunha et al. (33)	Empírico	0.4343

Continua na próxima página

Tabela Suplementar S3 – Comparação com literatura

Estação meteorológica	Artigo	Modelo de Aprendizado	R ²
CLAROS	(13)	MARS	0.720
MURIAE	Este trabalho	ELM	0.797
	Cunha et al. (33)	Empírico	0.5194
	Silva et al. (96)*	Empírico	0.7336
	(13)	MARS	0.864
OLIVEIRA	Este trabalho	ELM	0.704
OURO BRANCO	Este trabalho	ELM	0.647
	Cunha et al. (33)	Empírico	0.3425
	Silva et al. (96)*	Empírico	0.7336
	(13)	MARS	0.710
PARACATU	Este trabalho	ELM	0.587
PASSA QUATRO	Este trabalho	ELM	0.747
	Cunha et al. (33)	Empírico	0.3423
	Barbosa et al. (8)*	Empírico	0.72
	(13)	MARS	0.671
PASSOS	Este trabalho	ELM	0.749
	Cunha et al. (33)	Empírico	0.4364
	(13)	MARS	0.814
PATOS DE MINAS	Este trabalho	ELM	0.855
PATROCINIO	Este trabalho	ELM	0.735
	Cunha et al. (33)	Empírico	0.4846
	(13)	MARS	0.781
PIRAPORA	Este trabalho	ELM	0.683
RIO PARDO DE MINAS	Este trabalho	ELM	0.823
	Cunha et al. (33)	Empírico	0.5312
	Carvalho et al. (25)*	Empírico	0.711
	(13)	MARS	0.700
SACRAMENTO	Este trabalho	ELM	0.674
	Cunha et al. (33)	Empírico	0.4542
	(13)	MARS	0.744
SALINAS	Este trabalho	ELM	0.749
	Cunha et al. (33)	Empírico	0.4996
	Carvalho et al. (25)*	Empírico	0.711
	(13)	MARS	0.704
SÃO JOÃO DEL REI	Este trabalho	ELM	0.792
	Cunha et al. (33)	Empírico	0.5540
	Barbosa et al. (8)*	Empírico	0.72
	(13)	MARS	0.818
SÃO ROMÃO	Este trabalho	ELM	0.602
	Cunha et al. (33)	Empírico	0.4319
	Carvalho et al. (25)*	Empírico	0.711

Continua na próxima página

Tabela Suplementar S3 – Comparação com literatura

Estação meteorológica	Artigo	Modelo de Aprendizado	R ²
	(13)	MARS	0.585
SÃO SEBASTIÃO DO PARAISO	Este trabalho	ELM	0.683
	Este trabalho	ELM	0.833
SERRA	Cunha et al. (33)	Empírico	0.5476
DOS	Paula et al. (86)	Empírico	0.7272
AIMORES	(13)	MARS	0.808
SETE LAGOAS	Este trabalho	ELM	0.432
	Este trabalho	ELM	0.232
TEOFILO	Cunha et al. (33)	Empírico	0.5597
OTONI	Paula et al. (86)	Empírico	0.7272
	(13)	MARS	0.821
	Este trabalho	ELM	0.817
TIMOTEO	Cunha et al. (33)	Empírico	0.5398
	(13)	MARS	0.833
	Este trabalho	ELM	0.740
TRES	Cunha et al. (33)	Empírico	0.5175
MARIAS	(13)	MARS	0.721
UBERABA	Este trabalho	ELM	0.119
	Este trabalho	ELM	0.624
UBERLANDIA	Cunha et al. (33)	Empírico	0.3182
	(13)	ANN	0.613
	Este trabalho	ELM	0.674
	Cunha et al. (33)	Empírico	0.5022
UNAI	Silva et al. (97)*	Empírico	0.62
	Finzi et al. (51)*	Empírico	0.62
	(13)	MARS	0.729
	Este trabalho	ELM	0.761
VARGINHA	Cunha et al. (33)	Empírico	0.5385
	Barbosa et al. (8)*	Empírico	0.72
	(13)	MARS	0.836
	Este trabalho	ELM	0.846
VICOSA	Cunha et al. (33)	Empírico	0.5041
	Silva et al. (96)*	Empírico	0.7336
	(13)	MARS	0.829

Tabela Suplementar S4 – Utilização das cidades

Cidade	UF	Execução	1	2	3	4	5	6	7	8	9	10	Soma das Cidades
		R^2	0.698	0.718	0.751	0.681	0.709	0.692	0.747	0.670	0.780	0.707	
Soma das Execuções			66	70	62	66	63	73	80	68	66	61	
ABROLHOS	BA	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	8
AFONSO CLAUDIO	ES	×	✓	×	✓	×	×	×	×	✓	×	×	3
AGUAS EMENDADAS	DF	×	✓	×	×	✓	×	×	✓	✓	×	×	4
AGUAS VERMELHAS	MG	×	×	×	✓	×	×	✓	✓	×	×	✓	4
AIMORES	MG	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
ALEGRE	ES	×	×	×	✓	✓	✓	✓	✓	✓	✓	×	6
ALMENARA	MG	×	×	✓	✓	✓	✓	✓	✓	✓	×	✓	7
ANGRA DOS REIS	RJ	×	✓	✓	✓	✓	✓	×	✓	×	×	✓	6
ARACUAI	MG	×	✓	×	×	✓	✓	✓	×	×	×	×	3
ARAXA	MG	×	✓	×	×	✓	✓	✓	✓	✓	✓	×	6
ARIRANHA	SP	×	✓	×	×	✓	✓	✓	✓	✓	×	✓	6
BAMBUI	MG	×	✓	✓	✓	×	×	×	×	×	×	✓	4
BARBACENA	MG	×	✓	✓	×	×	×	✓	✓	×	×	✓	5
BARRETOS	SP	✓	×	✓	×	×	×	✓	×	✓	×	×	4
BELMONTE	BA	×	×	×	×	×	✓	✓	✓	✓	✓	×	5
BELO HORIZONTE	MG	×	×	✓	×	×	×	×	×	✓	✓	×	3
BELO HORIZONTE	MG	✓	✓	✓	✓	×	×	✓	×	×	✓	×	6
BERTIOGA	SP	✓	✓	✓	×	×	✓	×	✓	✓	×	×	6

Continua na próxima página

Tabela Suplementar S4 – Utilização das cidades

Cidade	UF	1	2	3	4	5	6	7	8	9	10	Soma
BOM JESUS DA LAPA	BA	✓	✓	✓	✓	✓	×	×	×	✓	✓	7
BRAGANCA PAULISTA	SP	✓	✓	✓	×	×	✓	✓	✓	✓	✓	8
BRASILIA	DF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
BRAZLANDIA	DF	✓	✓	×	✓	✓	✓	×	✓	×	✓	7
BRUMADO	BA	✓	×	✓	✓	✓	✓	×	×	✓	×	6
BURITIS	MG	×	✓	✓	×	×	×	✓	✓	×	×	4
CACHOEIRA PAULISTA	SP	×	×	✓	✓	×	×	✓	×	✓	✓	5
CALDAS	MG	×	×	×	×	✓	×	×	×	✓	×	2
CAMANDUCAIA	MG	×	×	×	×	×	×	✓	×	×	✓	2
CAMBUCI	RJ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
CAMPINA VERDE	MG	✓	×	×	✓	×	×	✓	✓	✓	×	5
CAMPOS DO JORDAO	SP	✓	✓	×	✓	✓	✓	×	✓	✓	✓	8
CAPELINHA	MG	×	×	×	×	✓	×	✓	×	×	×	2
CARATINGA	MG	✓	✓	✓	✓	✓	✓	×	×	✓	×	7
CARAVELAS	BA	✓	×	×	✓	✓	×	✓	✓	×	✓	6
CARMO	RJ	✓	✓	×	✓	×	×	×	✓	×	✓	5
CASA BRANCA	SP	✓	✓	×	×	✓	×	✓	✓	×	✓	6
CATALAO	GO	×	✓	✓	✓	×	×	×	✓	×	✓	5
CHAPADA GAUCHA	MG	×	×	✓	✓	×	×	✓	✓	×	×	4
CONCEICAO DAS ALAGOAS	MG	×	×	×	✓	×	✓	✓	×	×	✓	4
CORONEL PACHECO	MG	✓	×	×	✓	✓	×	×	×	✓	×	4
CORRENTINA	BA	×	×	×	×	×	×	×	✓	✓	✓	3

Continua na próxima página

Tabela Suplementar S4 – Utilização das cidades

Cidade	UF	1	2	3	4	5	6	7	8	9	10	Soma
CRISTALINA	GO	✓	×	×	×	×	×	×	✓	×	×	2
CRISTALINA	GO	×	×	×	×	✓	✓	×	×	✓	✓	4
CURVELO	MG	×	×	×	✓	×	✓	×	✓	×	×	3
DIAMANTINA	MG	✓	✓	✓	✓	×	×	×	✓	✓	×	6
DIVINOPOLIS	MG	×	✓	×	×	×	✓	✓	×	×	×	3
DORES DO INDAIA	MG	×	×	×	×	✓	✓	×	×	✓	×	3
ECOPORANGA	ES	×	✓	×	×	×	×	✓	✓	✓	✓	5
EDEIA	GO	×	×	✓	✓	×	×	✓	✓	×	×	4
ESPINOSA	MG	×	✓	✓	✓	×	×	✓	✓	✓	×	6
FLORESTAL	MG	×	×	×	×	✓	×	✓	✓	×	✓	4
FORMIGA	MG	×	✓	✓	✓	×	×	×	×	×	×	3
FRANCA	SP	✓	✓	✓	×	×	✓	✓	×	✓	✓	7
GOVERNADOR VALADARES	MG	×	×	✓	✓	✓	✓	✓	✓	×	✓	7
GUANAMBI	BA	×	×	✓	✓	✓	✓	✓	×	✓	×	6
GUANHAES	MG	×	✓	×	×	×	×	×	×	×	×	1
GUARDA-MOR	MG	×	×	×	✓	×	✓	✓	✓	×	×	4
IBIRITE	MG	✓	✓	×	×	×	✓	×	×	×	×	3
ITAMARAJU	BA	×	✓	×	×	×	✓	✓	×	×	×	3
ITAOBIM	MG	×	×	×	×	✓	×	✓	×	×	×	2
ITAPETINGA	BA	×	✓	✓	✓	×	×	×	×	✓	×	4
ITAPIRA	SP	✓	×	×	✓	✓	×	✓	✓	×	×	5
ITATIAIA	RJ	✓	×	×	×	✓	✓	×	×	✓	×	4

Continua na próxima página

Tabela Suplementar S4 – Utilização das cidades

Cidade	UF	1	2	3	4	5	6	7	8	9	10	Soma
ITUIUTABA	MG	×	×	×	×	×	×	✓	×	✓	×	2
ITUMBIARA	GO	✓	×	×	✓	✓	×	×	✓	✓	✓	6
ITUVERAVA	SP	✓	×	✓	✓	✓	✓	×	×	✓	✓	7
JALES	SP	×	✓	✓	×	×	×	×	×	✓	✓	4
JANAUBA	MG	✓	✓	✓	✓	✓	✓	×	×	✓	✓	8
JANUARIA	MG	×	✓	×	✓	×	✓	×	×	✓	×	4
JOAO PINHEIRO	MG	✓	✓	×	×	✓	×	×	×	✓	×	4
JUIZ DE FORA	MG	✓	×	×	×	×	✓	✓	×	×	×	3
LINHARES	ES	✓	×	✓	✓	×	✓	×	×	✓	✓	6
LUZIANIA	GO	×	×	×	×	×	✓	✓	×	×	×	2
MACHADO	MG	×	×	×	✓	×	×	×	✓	×	✓	3
MANHUACU	MG	×	✓	×	✓	×	✓	×	✓	✓	✓	6
MANTENA	MG	✓	✓	×	×	✓	×	×	×	✓	✓	5
MARIA DA FE	MG	×	✓	×	✓	✓	✓	×	×	✓	×	5
MARILANDIA	ES	✓	×	×	✓	✓	×	✓	✓	×	✓	6
MONTALVANIA	MG	✓	×	✓	×	✓	✓	×	×	✓	×	5
MONTES CLAROS	MG	✓	×	×	✓	✓	✓	✓	✓	×	✓	7
MORRINHOS	GO	×	✓	×	×	×	×	×	✓	×	✓	3
MURIAE	MG	×	×	×	×	×	×	×	×	✓	✓	2
NOVA FRIBURGO	RJ	×	✓	×	×	✓	✓	✓	✓	✓	✓	7
NOVA VENECIA	ES	✓	✓	✓	×	✓	×	×	✓	×	×	5
OLIVEIRA	MG	✓	×	×	×	×	✓	✓	×	×	×	3

Continua na próxima página

Tabela Suplementar S4 – Utilização das cidades

Cidade	UF	1	2	3	4	5	6	7	8	9	10	Soma
OURO BRANCO	MG	✓	✓	✓	×	×	✓	✓	×	×	✓	6
PARACATU	MG	×	✓	×	×	×	✓	✓	×	×	×	3
PARANAIBA	MS	×	✓	×	✓	×	✓	×	×	×	×	3
PARANOA (COOPA-DF)	DF	✓	×	×	×	✓	✓	×	×	×	×	3
PARATY	RJ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
PASSA QUATRO	MG	×	✓	✓	✓	×	✓	✓	✓	✓	×	7
PASSOS	MG	✓	✓	×	×	×	×	✓	×	×	×	3
PATOS DE MINAS	MG	×	✓	×	✓	✓	✓	×	✓	×	×	5
PATROCINIO	MG	×	✓	×	✓	×	×	✓	✓	✓	×	5
PETROPOLIS	RJ	✓	×	×	×	×	✓	×	✓	×	✓	4
PIRAPORA	MG	✓	×	✓	✓	✓	×	×	×	×	×	4
PIRES DO RIO	GO	×	×	×	×	×	✓	×	×	×	✓	2
PORTEIRINHA	MG	×	✓	×	×	×	✓	✓	×	×	×	3
PORTO SEGURO	BA	×	×	✓	×	✓	×	✓	✓	✓	✓	6
POSSE	GO	×	✓	×	✓	×	×	✓	✓	×	✓	5
PRADOPOLIS	SP	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
PRESIDENTE KENNEDY	ES	✓	×	×	×	✓	✓	✓	✓	✓	✓	7
RESENDE	RJ	×	×	✓	✓	×	✓	✓	×	✓	×	5
RIO CLARO	RJ	×	×	✓	✓	×	×	✓	×	×	×	3
RIO PARDO DE MINAS	MG	✓	×	✓	✓	×	✓	✓	×	✓	×	6
SACRAMENTO	MG	✓	✓	×	×	✓	×	✓	×	✓	×	5
SALINAS	MG	✓	✓	✓	×	×	×	✓	✓	✓	✓	7

Continua na próxima página

Tabela Suplementar S4 – Utilização das cidades

Cidade	UF	1	2	3	4	5	6	7	8	9	10	Soma
SANTA MARIA MADALENA	RJ	✓	×	✓	✓	✓	×	✓	✓	×	×	6
SANTA TERESA	ES	✓	✓	×	×	×	✓	✓	✓	✓	✓	7
SAO CARLOS	SP	✓	×	✓	×	×	×	✓	×	×	×	3
SAO JOAO DEL REI	MG	✓	×	✓	×	✓	✓	×	✓	×	✓	6
SAO LUIZ DO PARAITINGA	SP	✓	×	×	✓	✓	×	×	✓	✓	×	5
SAO MATEUS	ES	×	×	✓	✓	×	✓	✓	✓	×	✓	6
SAO PAULO	SP	×	×	✓	✓	×	✓	×	✓	✓	×	5
SAO ROMAO	MG	×	×	✓	×	✓	×	✓	×	✓	×	4
SAO SEBASTIAO DO PARAISO	MG	×	×	✓	✓	✓	×	×	×	×	✓	4
SERRA DOS AIMORES	MG	×	×	✓	×	✓	×	✓	✓	✓	✓	6
SETE LAGOAS	MG	✓	×	×	✓	✓	✓	✓	×	✓	✓	7
TAUBATE	SP	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10
TEOFILO OTONI	MG	✓	✓	×	×	×	×	✓	×	×	×	3
TERESOPOLIS	RJ	✓	×	✓	×	✓	✓	×	✓	✓	✓	7
TIMOTEO	MG	×	✓	×	×	×	✓	✓	×	✓	×	4
TRES MARIAS	MG	✓	✓	×	×	×	✓	✓	×	×	×	4
TRES RIOS	RJ	✓	✓	×	×	×	×	✓	✓	×	×	4
UBERABA	MG	×	×	×	×	×	×	✓	×	✓	×	2
UBERLANDIA	MG	×	×	✓	✓	×	✓	✓	✓	×	×	5
UNAI	MG	✓	×	✓	✓	✓	✓	✓	×	×	✓	7
VALENCA	RJ	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	9
VARGINHA	MG	✓	✓	✓	×	×	✓	✓	✓	×	×	6

Continua na próxima página

Tabela Suplementar S4 – Utilização das cidades

Cidade	UF	1	2	3	4	5	6	7	8	9	10	Soma
VENDA NOVA DO IMIGRANTE	ES	✓	✓	×	×	×	×	×	×	✓	×	3
VICOSA	MG	✓	✓	✓	×	×	×	✓	×	✓	✓	6
VILA VELHA	ES	×	✓	✓	×	×	✓	✓	×	×	×	4
VITORIA	ES	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	9
VITORIA DA CONQUISTA	BA	✓	✓	×	✓	✓	✓	✓	✓	×	×	7
VOTUPORANGA	SP	✓	✓	✓	×	×	✓	×	✓	✓	×	6

ANEXO A – Contribuições à literatura

Durante o desenvolvimento deste projeto os seguintes trabalhos foram publicados:

- (12) Samuel da Costa Alves Basílio, Rodrigo O. Silva, Camila M. Saporetti, and Leonardo Goliatt. Modeling global solar radiation using machine learning with model selection approach: A case study in tanzania. In Subarna Shakya, Klimis Ntalianis, and Khaled A. Kamel, editors, *Mobile Computing and Sustainable Informatics*, pages 155–168, Singapore, 2022. Springer Nature Singapore. ISBN 978-981-19-2069-1.
- (11) Samuel da Costa Alves Basílio, Camila Martins Saporetti, Zaher Mundher Yaseen, and Leonardo Goliatt. Global horizontal irradiance modeling from environmental inputs using machine learning with automatic model selection: A case study in tanzania. *Environmental Development*, page 100766, 2022. ISSN 2211-4645. doi: <https://doi.org/10.1016/j.envdev.2022.100766>.
URL <https://www.sciencedirect.com/science/article/pii/S2211464522000689>.
- (14) Samuel da Costa Alves Basílio, Camila M. Saporetti, and Leonardo Goliatt. An interdependent evolutionary machine learning model applied to global horizontal irradiance modeling. *Neural Computing and Applications*, 35(16):12099–12120, Jun 2023. ISSN 1433-3058. doi: 10.1007/s00521-023-08342-1.
URL <https://doi.org/10.1007/s00521-023-08342-1>.
- (13) Samuel da Costa Alves Basílio, Fernando Ferrari Putti, Angélica Carvalho Cunha, and Leonardo Goliatt. An evolutionary-assisted machine learning model for global solar radiation prediction in Minas Gerais region, southeastern Brazil. *Earth Science Informatics*, Apr 2023. ISSN 1865-0481. doi: 10.1007/s12145-023-00990-0. URL <https://doi.org/10.1007/s12145-023-00990-0>.
- (34) Samuel da Costa Alves Basílio, Cicero Manoel dos Santos, Alfeu Dias Martinho, and Leonardo Goliatt. Transfer learning for solar irradiation prediction in minas gerais, brazil. *Scientific Reports*, 15(1):40235, Nov 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-24095-4. URL <https://doi.org/10.1038/s41598-025-24095-4>.
- (10) Samuel Basilio, , and Leonardo Goliatt. An automl framework for transfer learning applied to solar radiation prediction: A case study in Juiz de Fora, Minas Gerais, Brazil. *XLVI CILAMCE*, 11 2025. doi: <https://doi.org/10.55592/cilamce2025.v5i.13251>.
- (9) Samuel Basilio and Leonardo Goliatt. Gradient boosting hybridized with exponential natural evolution strategies for estimating the strength of geopolymer self-compacting concrete. *Knowledge-Based Engineering and Sciences*, 3:1–16, 04 2022. doi: 10.51526/kbes.2022.3.1.1-16.

- (19) Tales Boratto, Camila Saporetti, Samuel Basilio, Alexandre Cury, and Leonardo Goliatt. Data-driven cymbal bronze alloy identification via evolutionary machine learning with automatic feature selection. *Journal of Intelligent Manufacturing*, 35: 1–17, 11 2022. doi: 10.1007/s10845-022-02047-3.