

MARCUS VINÍCIUS CONTES CALÇA

**DETECÇÃO DE OUTLIERS E PREENCHIMENTO DE LACUNAS EM MEDIDAS
DAS IRRADIÂNCIAS SOLARES ESPECTRAIS USANDO PRINCÍPIOS FÍSICOS E
TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL**

Botucatu

2022

MARCUS VINÍCIUS CONTES CALÇA

**DETECÇÃO DE OUTLIERS E PREENCHIMENTO DE LACUNAS EM MEDIDAS
DAS IRRADIÂNCIAS SOLARES ESPECTRAIS USANDO PRINCÍPIOS FÍSICOS E
TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL**

Tese apresentada à Faculdade de Ciências Agronômicas da UNESP - Câmpus de Botucatu, para obtenção do título de doutor em Agronomia (Energia na Agricultura).

Orientador: Alexandre Dal Pai.

Botucatu

2022

C144d	<p>Calça, Marcus Vinicius Contes</p> <p>Detecção de outliers e preenchimento de lacunas em medidas das irradiâncias solares espectrais usando princípios físicos e técnicas de inteligência artificial / Marcus Vinicius Contes Calça. -- Botucatu, 2022</p> <p>86 p. : il., tabs., fotos</p> <p>Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrônômicas, Botucatu</p> <p>Orientador: Alexandre Dal Pai</p> <p>1. Radiação solar espectral. 2. Detecção de medições errôneas. 3. Controle de qualidade. 4. Machine learning. 5. Modelagem. I. Título.</p>
-------	--

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

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DETECÇÃO DE OUTLIERS E PREENCHIMENTO DE LACUNAS EM MEDIDAS DAS IRRADIÂNCIAS SOLARES ESPECTRAIS USANDO PRINCÍPIOS FÍSICOS E TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

AUTOR: MARCUS VINÍCIUS CONTES CALÇA

ORIENTADOR: ALEXANDRE DAL PAI

Aprovado como parte das exigências para obtenção do Título de Doutor em Agronomia (Energia na Agricultura), pela Comissão Examinadora:

Prof. Dr. ALEXANDRE DAL PAI (Participação Presencial)
Bioprocessos e Biotecnologia / Faculdade de Ciências Agrômicas de Botucatu



Prof.ª Dr.ª VALERIA CRISTINA RODRIGUES SARNIGHAUSEN (Participação Presencial)
Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônômicas de Botucatu - UNESP



Prof. Dr. SERGIO AUGUSTO RODRIGUES (Participação Presencial)
Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônômicas de Botucatu - UNESP



Prof. Dr. FERNANDO DE LIMA CANEPPELE (Participação Virtual)
Engenharia de Biosistemas / Faculdade de Zootecnia e Engenharia de Alimentos - USP



Prof. Dr. OSVALDO CESAR PINHEIRO DE ALMEIDA (Participação Presencial)
Análise e Desenvolvimento de Sistemas / Faculdade de Tecnologia de Botucatu



Botucatu, 12 de dezembro de 2022

*A minha mãe Simônia A. de
Camargo Contes (in memoriam) e a
todas as pessoas que estiveram ao
meu lado nesta jornada,
dedico.*

AGRADECIMENTOS

Gostaria de agradecer, primeiramente, à Deus, por me proporcionar as oportunidades necessárias para que fosse possível realizar meus estudos. Minha eterna gratidão à minha mãe Simônia A. de Camargo Contes, que hoje descansa nos braços de Deus e que foi, sem dúvida alguma, minha maior apoiadora ao longo desta jornada, ao meu pai Valdeir P. Humberto Calça por todo o apoio, dedicação, investimentos e conselhos fornecidos. Agradeço também a minha avó Vera L. da Rocha Camargo, a minha irmã Michele M. Contes dos Santos e a minha companheira de vida Daniele C. Lopes Mariano, por estarem ao meu lado em momentos difíceis, pela compreensão, pelo suporte e companheirismo.

Deixo também meus agradecimentos ao professor Alexandre Dal Pai pelos conselhos e orientações concedidas ao longo dos anos e por sempre acreditar em meu potencial. Aos professores João Francisco Escobedo, Enzo Dal Pai, Sérgio Augusto Rodrigues, Carlos Roberto Pereira Padovani, Fernando de Lima Caneppele, Osvaldo Cesar Pinheiro de Almeida e à professora Valéria Cristina Rodrigues Sarnighausen, deixo meu muito obrigado por todos os sábios ensinamentos. Aos colaboradores do Departamento de Bioprocessos e Biotecnologia Firmo Sousa Campos, Marcelo Felipe Matias, José Eduardo Gomes Montanha e Raquel Francisco Rubio expresso minha gratidão pela amizade e apoio no desenvolvimento dessa pesquisa.

Saliento minha gratidão a todo o corpo docente do Programa de Pós-Graduação em Agronomia - Energia na Agricultura, aos colaboradores(as) da Seção Técnica de Pós-Graduação, da Biblioteca, da Manutenção e da limpeza. Aos amigos e colegas de pesquisa, Matheus Rodrigues Raniero, José Rafael Franco, Marcos Roberto R. Bica, Domingos Mário Zeca Fernando e Luiz Eduardo V. Montanha expresso meus agradecimentos por se prestarem a me ouvir, aconselhar e apoiar no decorrer dessa experiência, tornando-a mais expressiva e valorosa.

O presente estudo foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

A radiação solar de ondas curtas que incide sob a superfície terrestre é composta espectralmente pelas componentes ultravioleta (290 a 400nm), visível (400 a 700nm) e infravermelha (700 a 4000nm). O conhecimento sobre a sua disponibilidade na superfície terrestre permite que órgãos públicos e instituições acadêmicas planejem adequadamente projetos energéticos no Brasil. Para isso, se torna necessário a implantação de sistemas terrestres de medição, que compreendem, naturalmente, valores errôneos causados por adversidades técnicas e operacionais. Neste sentido, o objetivo deste estudo foi implementar um procedimento de controle de qualidade, para identificar valores errôneos, e outro para preencher lacunas das irradiâncias solares ultravioleta, visível e infravermelha sub-horárias (W/m^2), coletadas em Botucatu (SP) - Brasil no período de 2001 a 2006. O procedimento de controle de qualidade foi criado em função de três princípios físicos e um intervalo de confiança estatística. Em sequência, o procedimento de preenchimento de lacunas foi implementado em função de um modelo de Machine Learning e avaliado a partir do coeficiente de determinação (R^2), erro médio absoluto (MAE e rMAE), e da raiz do erro quadrático médio (RMSE e rRMSE). Foi possível detectar a partir do procedimento de controle de qualidade 1,003%, 0,989% e 0,806% de medições errôneas e isoladas das irradiâncias solares ultravioleta, infravermelha e visível, respectivamente, com relação aos totais medidos. O modelo de preenchimento de lacunas obteve métricas, com relação a sua capacidade de estimar valores, de R^2 de 0,99565, rMAE de 2,849% e rRMSE de 4,821% para a irradiância solar infravermelha, bem como R^2 de 0,99562, rMAE de 2,886% e rRMSE de 4,875% para a visível e, por fim, R^2 de 0,99137, rMAE de 3,923% e rRMSE de 6,889% para a ultravioleta. A partir deste estudo, foi possível avaliar que o primeiro e o terceiro princípio físico foram o mais e o menos restritivos do controle de qualidade, respectivamente, ao detectar medidas errôneas e isoladas, consideradas outliers por se afastarem consideravelmente das demais. Constatou-se, também, que os modelos de Machine Learning obtiveram resultados semelhantes as modelagens de Escobedo et al. (2011) para resoluções temporais horárias e diárias, o que possibilita utilizá-los para preencher lacunas das irradiâncias solares espectrais.

Palavras-Chave: radiação solar espectral; detecção de medições errôneas; controle de qualidade; machine learning; modelagem.

ABSTRACT

The shortwave solar radiation that falls on the Earth's surface is spectrally composed of ultraviolet (290 to 400nm), photosynthetically active (400 to 700nm) and near-infrared (700 to 4000nm) components. The knowledge about its availability on the Earth's surface allows public agencies and academic institutions to plan energy projects in Brazil. For this, it is necessary to implement terrestrial measurement systems, which naturally include erroneous values caused by technical and operational adversities. In this sense, the aim of this study was to implement a quality control procedure, to identify erroneous values, and to fill gaps in sub hourly ultraviolet, photosynthetically active, and near-infrared solar irradiance (Wm^{-2}), collected in Botucatu (SP) - Brazil from 2001 to 2006. The quality control procedure was created based on three physical principles and a statistical confidence interval. The gap-filling procedure was implemented based on a Machine Learning model and evaluated based on the coefficient of determination (R^2), mean absolute error (MAE and rMAE), and the root mean squared error (RMSE and rRMSE). It was possible to detect from the quality control procedure 1.003%, 0.989% and 0.806% of erroneous and isolated measurements of ultraviolet, near infrared and photosynthetically active solar irradiances, respectively, in relation to the total measured. The gap-filling model obtained metrics, regarding its ability to estimate values, of R^2 of 0.99565, rMAE of 2.849% and rRMSE of 4.821% for near infrared, as well as R^2 of 0.99562, rMAE of 2.886% and rRMSE of 4.875% for the photosynthetically active and R^2 of 0.99137, rMAE of 3.923% and rRMSE of 6.889% for the ultraviolet solar irradiances. From this study it was possible to assess that the first physical principle was the most restrictive, and the third physical principle the least restrictive, of quality control, when detecting erroneous measurements. All measurements flagged as erroneous or isolated are considered outliers because they are considerably different from the others. Finally, it was found that the gap-filling models obtained results similar to the models by Escobedo et al. (2011) for hourly and daily temporal resolutions, which makes it possible to use them to estimate measurements of spectral solar irradiance.

Keywords: spectral solar radiation; detection of erroneous measurements; quality control; machine learning; modeling.

LISTA DE ILUSTRAÇÕES

Figura 1 - Espectro eletromagnético da radiação solar	24
Figura 2 - Componentes da radiação solar	26
Figura 3 - Distribuição espectral da irradiância solar na superfície terrestre.	27
Figura 4 - Aspectos técnicos do pireliômetro.....	28
Figura 5 - Aspectos técnicos do piranômetro	29
Figura 6 - Aspectos técnicos e composição do piranômetro de fotodiodo.....	30
Figura 7 - Fatores que afetam as medidas da irradiância solar.....	32
Figura 8 - Série de medidas da Estação Meteorológica Lageado (2022).....	38
Figura 9 - Radiômetros do Laboratório de Agrometeorologia e Radiometria Solar ..	39
Figura 10 - Dispersão das irradiâncias solares espectrais em estado bruto	56
Figura 11 - Percentual de medidas invalidadas pelos princípios físicos.....	61
Figura 12 - Percentual de medidas invalidadas pelo intervalo estatístico	65
Figura 13 - Dispersão das irradiâncias solares espectrais após o procedimento.....	69
Figura 14 - Estrutura do algoritmo de Gradient Boosting (Machine Learning).....	72
Figura 15 - Resultado do modelo de ML da irradiância solar infravermelha.....	74
Figura 16 - Resultado do modelo de ML da irradiância solar visível	75
Figura 17 - Resultado do modelo de ML da irradiância solar ultravioleta	76

LISTA DE TABELAS

Tabela 1 - Instrumentos de medição das irradiâncias solares espectrais.....	40
Tabela 2 - Diagnósticos do procedimento de controle de qualidade.....	41
Tabela 3 - Constante solar das irradiâncias solares	43
Tabela 4 - Classificação da cobertura de céu em função do KT.....	45
Tabela 5 - Intervalos estatísticos de acordo com o nível de confiança	46
Tabela 6 - Recursos de previsão dos modelos de Machine Learning.....	49
Tabela 7 - Quantidade total de medidas sem validação	53
Tabela 8 - Quantidade de campos nulos na série de medidas	54
Tabela 9 - Quantidade total de medidas não-nulas sem validação.....	54
Tabela 10 - Quantidade de medidas invalidadas pelos limites possíveis	57
Tabela 11 - Quantidade de medidas invalidadas pelos limites raros	58
Tabela 12 - Quantidade de medidas invalidadas pelo topo da atmosfera	59
Tabela 13 - Estatísticas descritivas das irradiâncias solares espectrais.....	62
Tabela 14 - Quantidade de medidas invalidadas pelos intervalos estatísticos	64
Tabela 15 - Quantidade de medidas invalidadas por ano	66
Tabela 16 - Quantidade final de valores sinalizados como inválidos	67

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica (BR)
ARIMA	Autoregressive Integrated Moving Average
BSRN	Rede de Radiação em Superfície Terrestre (CH)
CIE	Comissão Internacional de Iluminação (FR)
LSTM	Long Short Term Memory
Cos	Cosseno
D	Dia do Ano
E	Ângulo de Elevação Solar
Exc	Fator de Correção da Excentricidade da Órbita Terrestre
H	Hidrogênio
He	Hélio
Hd	Hora e Décimo de Hora
INV	Medida Inválida
IV	Radiação Solar Infravermelha
IV ₀	Radiação Solar Infravermelha no Topo da Atmosfera
G ₀ ou IE	Radiação Solar Global no Topo da Atmosfera
G ou IG	Radiação Solar Global
G _{dif}	Radiação Solar Difusa
G _{dir}	Radiação Solar Direta
G _n	Radiação Solar Direta em Incidência Normal
I _{sc}	Constante Solar
K	Temperatura em Kelvin
KT	Índice de Transmissividade Atmosférica
M	Metros
MAE	Mean Absolute Error
Mm	Milímetros
Ms	Milisegundos
N	Fotoperíodo
n	Horas de Briho Solar
Nm	Nanometro
NREL	Laboratório Nacional de Energias Renováveis (EUA)

O	Oeste
PAR	Radiação Solar Visível
PAR ₀	Radiação Solar Visível no Topo da Atmosfera
PSP	Piranômetro Espectral de Precisão
RMIB	Instituto Meteorológico Real da Bélgica (BE)
RMSE	Root Mean Square Error
R ²	Coeficiente de Determinação
S	Sul
s	Segundos
SGBDR	Sistema de Gerenciamento de Banco de Dados Relacional
SQL	Structured Query Language
UV	Radiação Solar Ultravioleta
UV-A	Faixa A da Radiação Solar Ultravioleta
UV-B	Faixa B da Radiação Solar Ultravioleta
UV-C	Faixa C da Radiação Solar Ultravioleta
UV ₀	Radiação Solar Ultravioleta no Topo da Atmosfera
VAL	Medida Válida
Z	Ângulo Zenital
W	Watt
WMO	Organização Internacional de Meteorologia (CH)
WRMC	Centro Mundial de Monitoramento de Radiação (CH)
3D-CNN	3D Convolutional Neural Network
μV	Microvolts
λ	Comprimento de Onda
δ	Declinação Solar
Φ	Latitude Geográfica
°C	Graus Celsius
Ω	Ângulo Horário
%	Quantidade Percentual de Medidas
μ	Média
σ	Desvio Padrão

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Justificativa.....	22
1.2	Objetivo Geral.....	22
1.3	Objetivos Específicos	23
2	REVISÃO DE LITERATURA	24
2.1	Radiação Solar	24
2.2	Componentes Espectrais da Radiação Solar	25
2.3	Medição da Radiação Solar	27
2.4	Controle de Qualidade da Radiação Solar	30
2.5	Preenchimento de Lacunas de Radiação Solar	34
3	MATERIAL E MÉTODOS	37
3.1	Localização e Clima	37
3.2	Instrumentação e Medidas	38
3.3	Controle de Qualidade	41
3.3.1	Princípios Físicos de Detecção de Medidas Errôneas.....	42
3.3.2	Intervalo Estatístico de Detecção de Medidas Isoladas.....	44
3.4	Preenchimento de Lacunas com Machine Learning	47
4	RESULTADOS E DISCUSSÃO	53
4.1	Análise Exploratória das Irradiâncias Solares.....	53
4.2	Princípios Físicos de Detecção de Medidas Errôneas.....	57
4.3	Intervalo Estatístico de Detecção de Medidas Isoladas.....	62
4.4	Preenchimento de Lacunas com Machine Learning	70
5	CONCLUSÕES	78
	REFERÊNCIAS.....	81

1 INTRODUÇÃO

O conhecimento sobre o potencial disponível de irradiância solar na superfície terrestre é imprescindível para órgãos públicos, instituições acadêmicas e organizações não governamentais desenvolverem projetos baseados em energias renováveis mais adequadamente no Brasil (PEREIRA et al., 2017). Esse conhecimento é gerado a partir de medições obtidas na superfície terrestre, e permite avaliar o potencial solar para o aproveitamento térmico (energia termossolar), dimensionamento de sistemas distribuídos e de plantas de energia solar fotovoltaica, capacidade de geração de biomassa para a produção de eletricidade, além do aquecimento passivo e ativo em centros urbanos ou em ambientes rurais (PASHIARDIS; KALOGIROU, 2016; YOUNES; CLAYWELL; MUNEEER, 2005).

O fluxo de irradiância solar que atinge um determinado local do planeta Terra, normalmente, é mensurado por meio do uso de piranômetros ou pireliômetros e pode ser classificado como direto/direta, quando incide diretamente do disco solar, difuso/difusa, quando sofre influência dos fenômenos físicos de reflexão ou espalhamento na atmosfera, e global/total, quando é a soma dos dois anteriores e representa a densidade total de energia solar incidente sob uma determinada área. Essas irradiâncias solares possuem três componentes espectrais classificadas de acordo com seu comprimento de onda, a ultravioleta, com intervalo espectral de 290 a 385 ou 400nm, a visível, com intervalo espectral de 400 a 700nm, e a infravermelha, com intervalo espectral de 700 a 4000nm (ESCOBEDO et al., 2011).

O piranômetro é o instrumento constituído por um sensor de termopilha ou de fotodiodo, usado para medir a irradiância solar global e difusa. O pireliômetro é o instrumento composto por um sensor de termopilha, que acoplado a um rastreador solar permite medir a irradiância solar direta (PEREIRA et al., 2017). Apesar da evolução tecnológica desses instrumentos, medir a irradiância solar não é uma tarefa fácil, em virtude das incertezas contidas no processo, que afetam diretamente os valores captados. Essas incertezas podem ser classificadas como instrumentais, quando derivam do funcionamento do sensor, ou operacionais, quando são causadas por manutenção precária ou interferências externas (YOUNES; CLAYWELL; MUNEEER, 2005; MUNEEER; YOUNES; MUNAWWAR, 2007; MORADI, 2009).

1.1 Justificativa

A irradiância solar ultravioleta vem sendo, constantemente, aplicada na decomposição residual de águas por meio de concentradores solares, porém também tem sido estudada por afetar nocivamente a vida humana, pois causa câncer de pele e catarata ocular, assim como por modificar a fisiologia vegetal (ROSSI, et al., 2018). A irradiância solar visível, também conhecida como fotossinteticamente ativa, é usada pelos processos bioquímicos das plantas para fazer a fotossíntese, assim como tem considerável importância em processos agrícolas (ESCOBEDO et al., 2009). A irradiância solar infravermelha é utilizada por satélites em estudos de nuvens na atmosfera terrestre, bem como causa um aumento de temperatura em ambientes protegidos e contribui no crescimento vegetativo (LAMNATOU; CHEMISANA, 2013).

Nesta perspectiva, é necessário o conhecimento, adquirido a partir de medições climáticas com ampla cobertura temporal e espacial, a fim de embasar estudos e investigações científicas. No entanto, medir parâmetros climáticos, principalmente a irradiância solar, não é uma tarefa fácil, por conta das inconsistências geradas pelas incertezas do sistema de medição (CALÇA et al., 2019). Essas incertezas podem ser reduzidas e as inconsistências identificadas e quantificadas a partir de processos de controle de qualidade e de validação de medidas. Que possuem como intuito principal identificar minuciosamente, a partir de algoritmos computacionais e regras de avaliação, valores errôneos e sinalizá-los, para que assim possa-se gerar maior confiança acerca do que foi mensurado.

1.2 Objetivo Geral

O objetivo desta pesquisa foi desenvolver um procedimento de controle de qualidade e de preenchimento de lacunas para medições da irradiância solar ultravioleta, visível e infravermelha em superfície terrestre horizontal, considerando a escala de tempo sub-horária (5 minutos em W/m^2), por meio da aplicação de algoritmos computacionais e modelos de Machine Learning (inteligência artificial). Tal que, permitam a identificação de valores fisicamente impossíveis, extremamente raros, maiores que o limite teórico pré-definido e fora do intervalo estatístico de confiança para as componentes das irradiâncias solares espectrais, de modo a sinalizá-los como inválidos e gerar documentos quantitativos, assim como possibilitem o preenchimento de medições ausentes (lacunas) usando valores modelados.

1.3 Objetivos Específicos

- a)** Implementar um algoritmo para verificar se as medições (5 minutos em W/m^2) das irradiâncias solares ultravioleta, visível e infravermelha estão dentro dos limites fisicamente possíveis.
- b)** Implementar um algoritmo para verificar se as medições (5 minutos em W/m^2) das irradiâncias solares ultravioleta, visível e infravermelha estão dentro dos limites para serem consideradas extremamente raras.
- c)** Implementar um algoritmo para comparar os valores medidos (5 minutos em W/m^2) das irradiâncias solares ultravioleta, visível e infravermelha com as estimativas obtidas para o topo da atmosfera.
- d)** Implementar um algoritmo para verificar se as medições (5 minutos em W/m^2) das irradiâncias solares ultravioleta, visível e infravermelha estão dentro dos intervalos estatísticos de confiança em função da cobertura de céu.
- e)** Implementar um modelo de previsão de valores (5 minutos em W/m^2) e preenchimento de lacunas das irradiâncias solares ultravioleta, visível e infravermelha usando algoritmos de Machine Learning.

2 REVISÃO DE LITERATURA

2.1 Radiação Solar

Radiação solar é o termo genérico atribuído a energia do Sol que incide sobre a superfície terrestre por meio de movimentos ondulatórios, derivados de um fenômeno de natureza eletromagnética. É a energia que controla a dinâmica da atmosfera, o clima (temperatura e umidade do ar, velocidade e direção do vento, pressão atmosférica e chuva) e outras fontes energéticas do planeta Terra. Sua atividade é decorrente das reações de fusão nuclear dos átomos de hidrogênio (H) para formar o hélio (He). Quando interceptada pela atmosfera terrestre, a radiação solar tem seus diferentes comprimentos de onda (λ) interagindo de formas distintas com os componentes atmosféricos (gases e aerossóis) através dos processos físicos de reflexão, absorção e espalhamento (IQBAL, 1983; PEREIRA et al., 2017).

O comprimento de onda, distância que separa duas cristas consecutivas, é o que caracteriza a radiação solar de acordo com a óptica ondulatória. Seu domínio cobre uma ampla parcela do espectro eletromagnético (Figura 1), desde a faixa do ultravioleta ao infravermelho. Cerca de 81% da energia solar que chega ao sistema Terra/Atmosfera abrange a região do visível ao infravermelho próximo (VAREJÃO-SILVA, 2006; PEREIRA et al., 2017). Neste sentido, ao incidir sob a superfície terrestre a radiação solar pode ser classificada em três componentes espectrais de acordo com seu comprimento de onda, a ultravioleta, com intervalo espectral de 290 a 385 ou 400nm, a fotossinteticamente ativa, com intervalo espectral de 400 a 700nm, e a infravermelha, com intervalo espectral de 700 a 4000nm (ESCOBEDO et al., 2011).

Figura 1 - Espectro eletromagnético da radiação solar.



Fonte: Pereira et al., 2017 (p. 15).

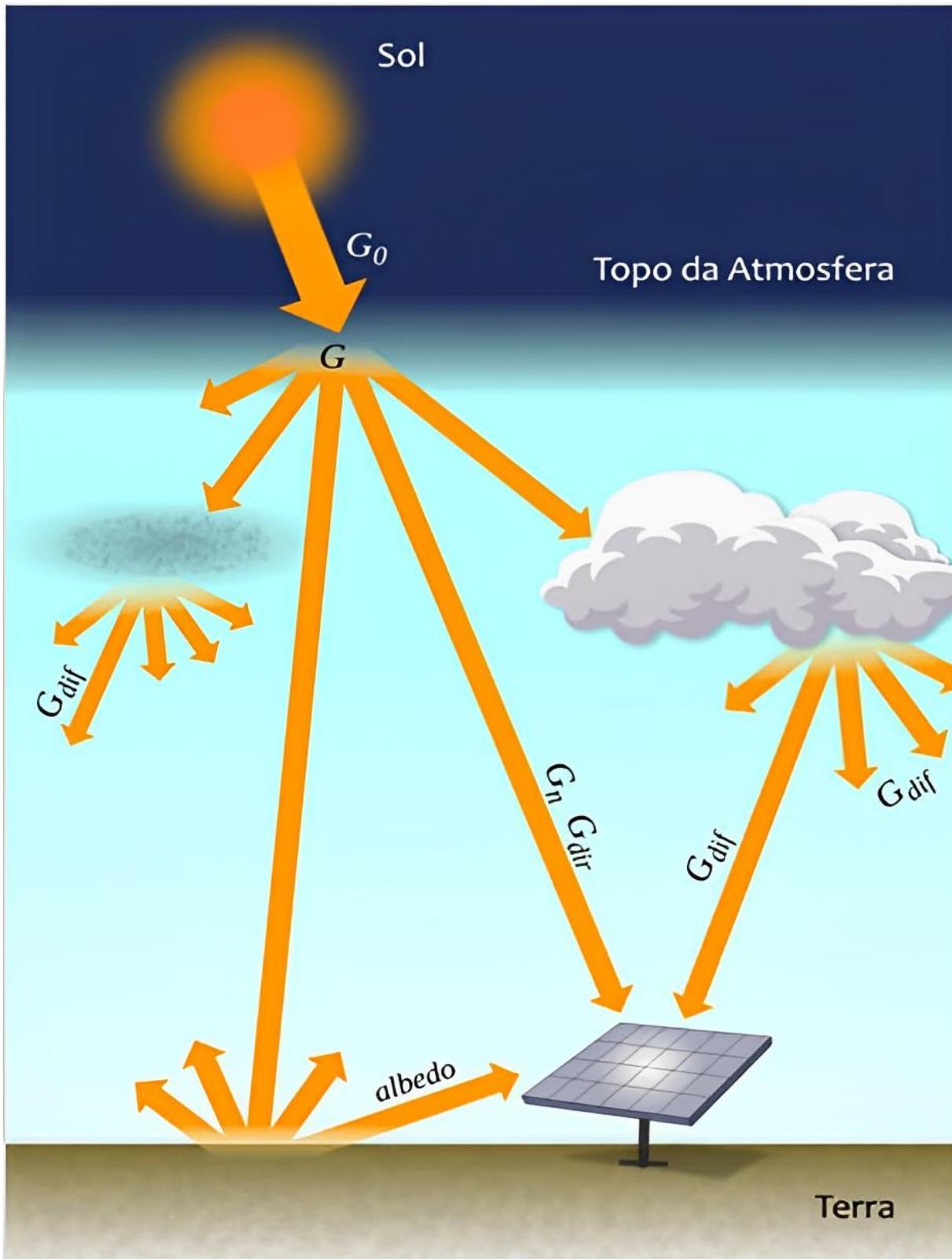
A radiação solar tem grande importância para a fotossíntese vegetal, ciclos hidrológicos, refletividade de superfícies, dinâmica da atmosfera e dos oceanos, iluminação passiva e ativa de ambientes, processos fotoquímicos e para o desenvolvimento energético sustentável na geração de energia elétrica sem causar impactos ambientais ao planeta Terra (IQBAL, 1983; VAREJÃO-SILVA, 2006;

PEREIRA et al., 2017). Além de sua importância nos processos já citados, deve-se destacar a relevância que as componentes espectrais da radiação solar (ultravioleta, visível e infravermelha) tem para estudos sobre a decomposição de águas residuais, crescimento de culturas agrícolas e composição atmosférica de nuvens (ROSSI et al., 2018; ESCOBEDO et al., 2009; LAMNATOU; CHEMISANA, 2013).

2.2 Componentes Espectrais da Radiação Solar

A radiação oriunda do Sol disponível em um plano imaginário no topo da atmosfera terrestre é conhecida como radiação solar extraterrestre (G_0). Ao atravessar a atmosfera e, em um determinado momento, atingir uma superfície, é conhecida como radiação solar global (G), e pode ser particionada em (Figura 2) radiação solar direta (G_{dir} e G_n) e radiação solar difusa (G_{dif}). A direta é a parcela da radiação solar global proveniente diretamente do disco solar, quando ele se mostra visível, e que não sofreu os processos de absorção ou espalhamento na atmosfera. A difusa é a parcela da radiação solar global que resultou da ação de espalhamento pelas nuvens, gases e partículas atmosféricas, assim como que sofreu desvios em seu percurso até atingir a área de interesse (VAREJÃO-SILVA, 2006; WMO, 2012; PEREIRA et al., 2017).

Figura 2 - Componentes da radiação solar.

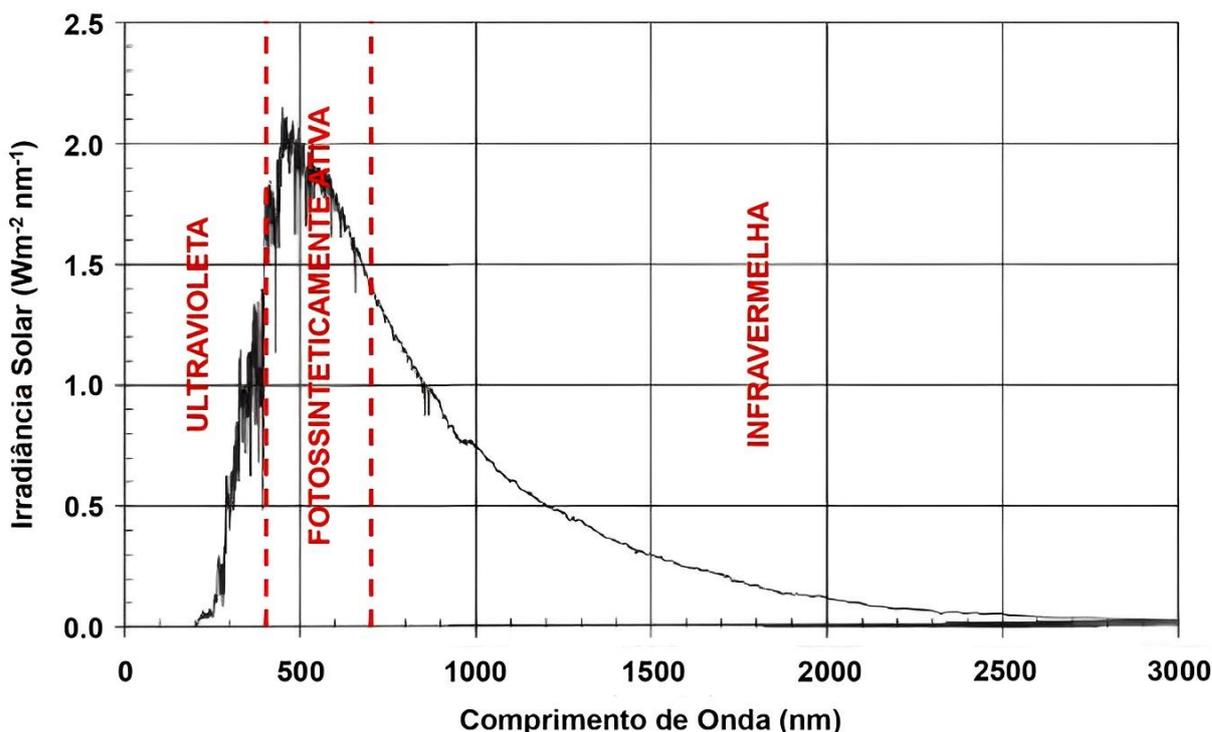


Fonte: Pereira et al., 2017 (p. 19).

Com relação a radiação solar espectral (Figura 3), a ultravioleta pode ser decomposta em outras três componentes, a UV-A (faixa espectral de 315 a 400nm) e

UV-B (faixa espectral de 280 a 315nm), que causam efeitos danosos em sistemas biológicos (doenças de pele e catarata ocular) e geram a foto-decomposição de águas residuais e de contaminantes industriais, e a UV-C (faixa espectral de 100 a 280nm), que é absorvida pela camada de ozônio e outros constituintes atmosféricos (CORRÊA, 2011; ESCOBEDO et al., 2011). A radiação solar fotossinteticamente ativa, ou visível (faixa espectral de 400 a 700nm), tem grande importância para o processo bioquímico de fotossíntese vegetal (transformação de água e dióxido de carbono em glicose), sendo de extrema importância, portanto, no que se refere a pesquisas relacionadas a processos agrícolas (VIGNOLA; MICHALSKY; STOFFEL, 2012). Por fim, a radiação solar infravermelha próxima (faixa espectral de 700 a 4000nm), pode ser utilizada em estudos sobre a composição e o vapor d'água presente nas nuvens, bem como sobre seu impacto na temperatura interna de estufas, que gera um aumento na produção agrícola (ROSSI et al., 2018).

Figura 3 - Distribuição espectral da irradiância solar na superfície terrestre.



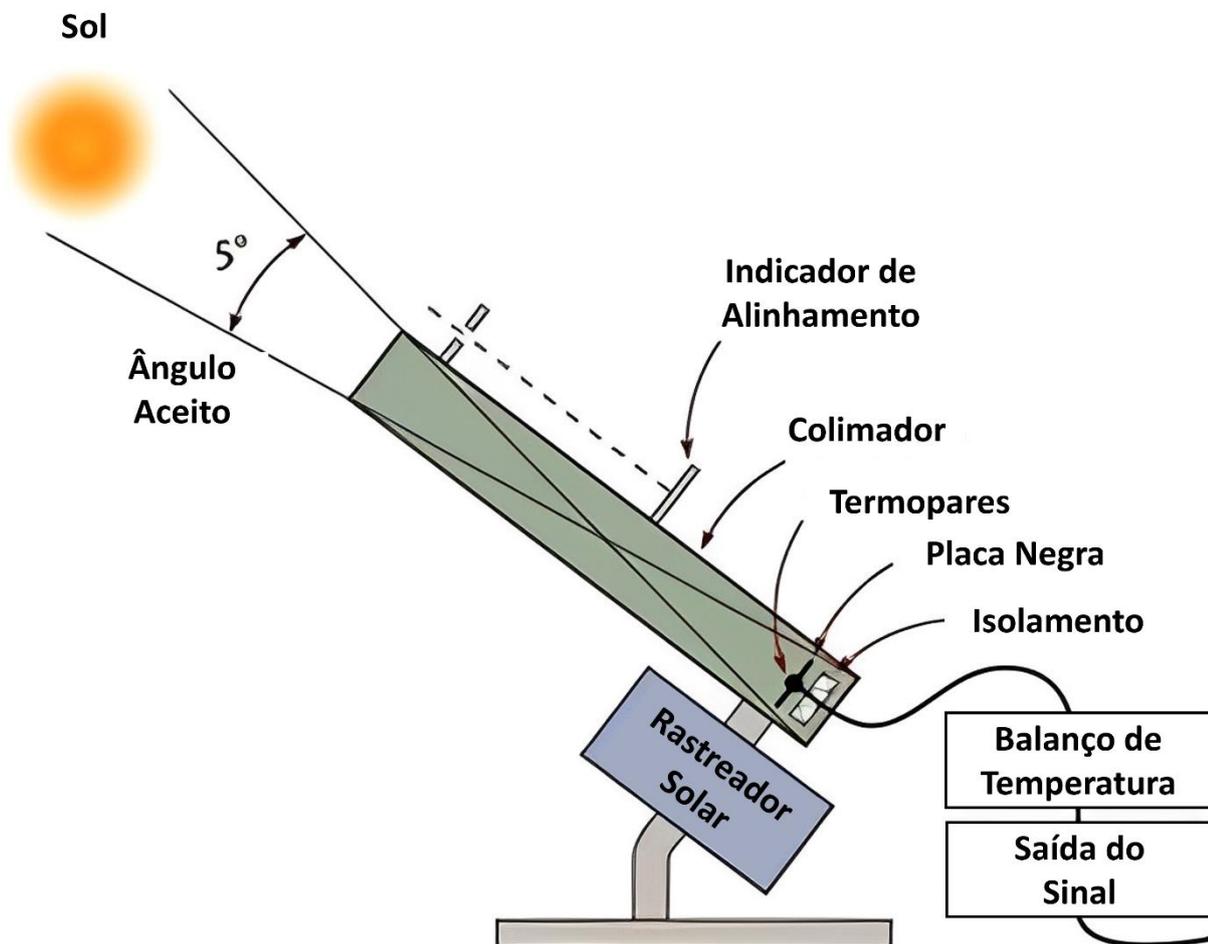
Fonte: Adaptado de Vignola; Michalsky; Stoffel, 2012 (p. 26).

2.3 Medição da Radiação Solar

A radiação solar tem sua variação de acordo com os movimentos do planeta Terra e pode ser afetada pelos seus constituintes atmosféricos (nuvens, vapor d'água,

gases e aerossóis). Isto requer que a instrumentação para aquisição de suas medidas tenha sensibilidade para detectar todas estas variabilidades (SHI, et al., 2008; PEREIRA et al., 2017). A radiação solar direta (G_{dir} e G_n) é obtida por meio de um pireliômetro (Figura 4), que possui uma estreita abertura para coletar apenas a radiação oriunda diretamente do disco solar, com uma pequena contribuição da radiação circumsolar (região vizinha do disco solar). O pireliômetro é um aparelho que deve ser apontado perpendicularmente ao Sol, acompanhando seu movimento ao longo do dia (GUEYMARD; MYERS, 2008; PEREIRA et al., 2017).

Figura 4 - Aspectos técnicos do pireliômetro.

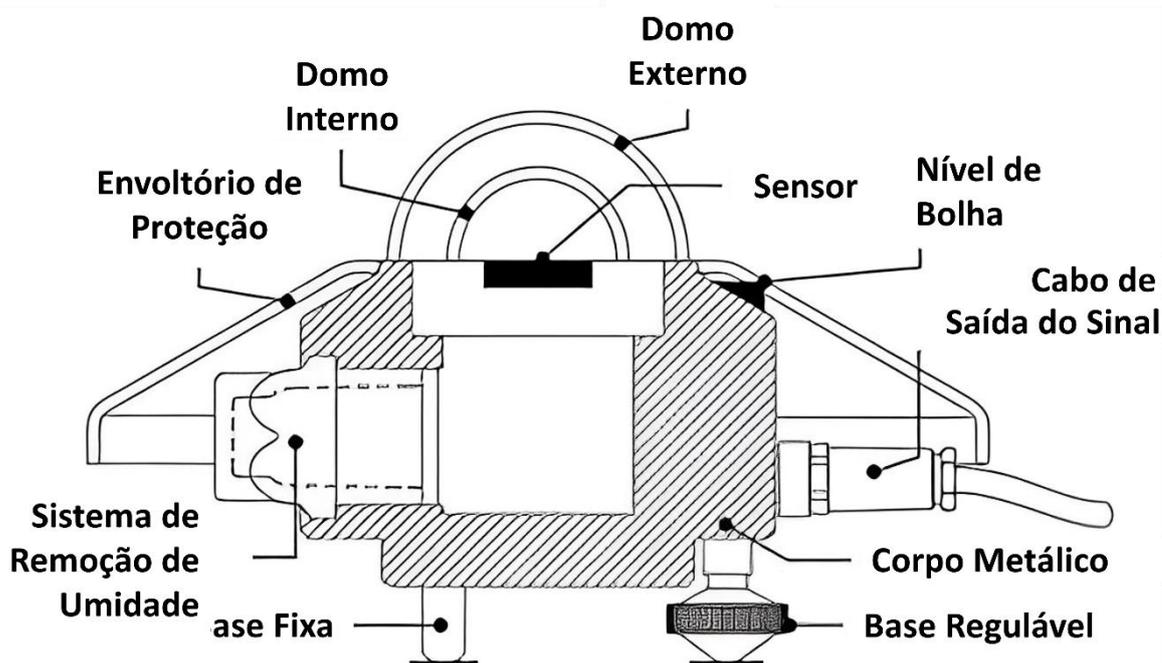


Fonte: Kipp & Zonen (2016) apud Pereira et al. (2017).

A radiação solar global (G) e a difusa (G_{dif}) são obtidas por meio de um piranômetro (Figura 5), que tem um campo de visão de 180° e geralmente cobre uma faixa espectral de 250nm a 3000nm (DRIEMEL et al., 2018). Esse instrumento é utilizado para medir a radiação solar por meio de um sensor de termopilha, que

converte a energia térmica em elétrica, ou de fotodiodo (silício), que funciona por meio do efeito fotoelétrico. O piranômetro responsável por captar a componente difusa da radiação solar deve ser utilizado junto de um método de sombreamento. Esse método é responsável pela supressão da incidência do feixe solar (radiação solar direta) no sensor e obtenção da radiação solar oriunda dos fenômenos de espalhamento e reflexão (GUEYMARD; MYERS, 2008; PEREIRA et al., 2017).

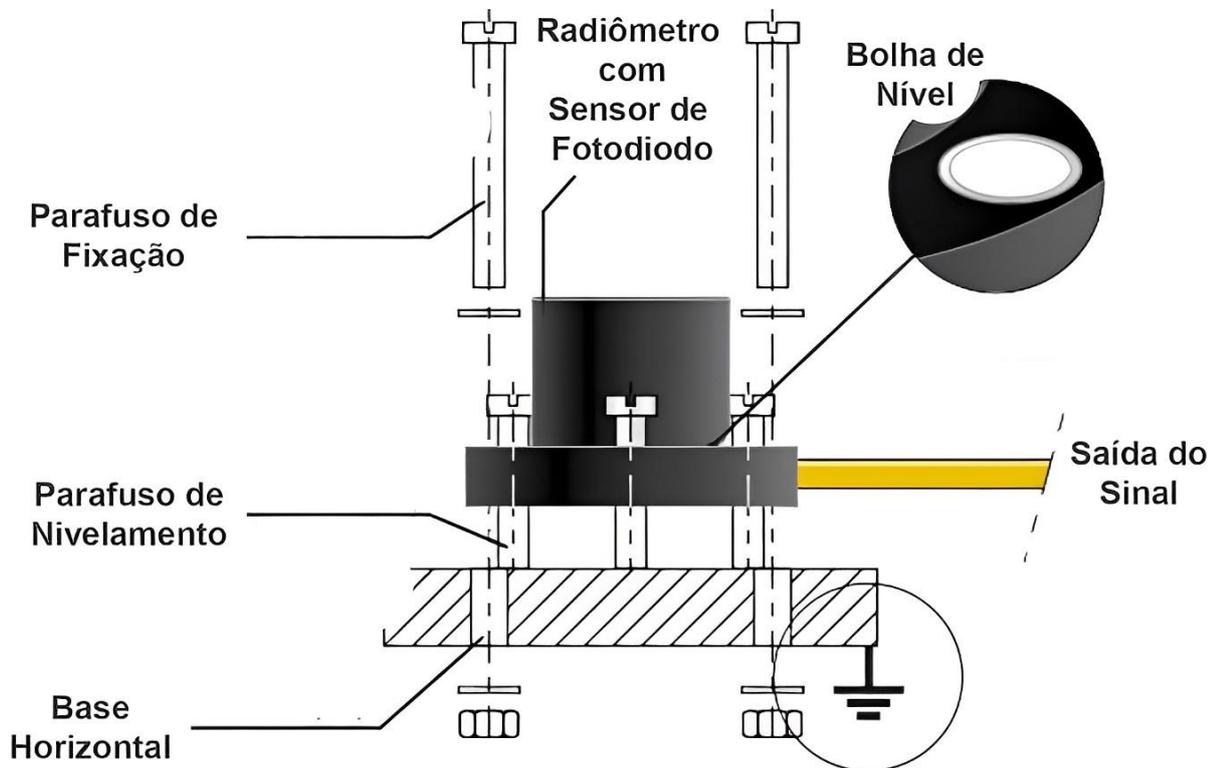
Figura 5 - Aspectos técnicos do piranômetro.



Fonte: Kipp & Zonen (2016) apud Pereira et al. (2017).

As radiações solares ultravioleta e infravermelha podem ser obtidas, similarmente, a partir de piranômetros de termopilha (Figura 6), chamados também de radiômetros, com uma modificação na cúpula, que possuem por sua vez, um filtro de seleção espectral da luz sob o elemento sensível, a fim de detectar apenas os respectivos comprimentos de onda dessas componentes (290 a 400nm e 700 a 4000nm, respectivamente). De forma que, a radiação solar fotossinteticamente ativa, ou visível, é medida por meio de um radiômetro composto por um sensor de fotodiodo de silício (Figura 6), também com filtros de seleção espectral, que aceitam emissões de luz no intervalo espectral de 400 a 700nm (VIGNOLA; MICHALSKY; STOFFEL, 2012; KIPP & ZONEN, 2022a; KIPP & ZONEN, 2022b).

Figura 6 - Aspectos técnicos e composição do piranômetro de fotodiodo de silício.



Fonte: Adaptado de Kipp & Zonen (2022).

Apesar da constante evolução tecnológica na construção de sensores de fotodiodo de silício e de termopilhas, medir corretamente a irradiância solar incidente na superfície da Terra é mais difícil do que outras variáveis climatológicas. Isto se dá devido a ocorrência de falhas técnicas, com origem instrumental, e de manutenção precária, com origem operacional, no sistema de monitoramento (YOUNES; CLAYWELL; MUNEEER, 2005; MORADI, 2009). Bem como, existem custos de operação (treinamento de pessoal especializado) que acarretam mais dificuldades ao processo. De modo que se necessita, então, de um procedimento de controle de qualidade das medições para detectar valores errôneos, que derivam dessas adversidades, para corrigir imediatamente os problemas que os causam.

2.4 Controle de Qualidade da Radiação Solar

O controle de qualidade é um processo composto por procedimentos operacionais (calibração de instrumentos, inspeções, manutenção preventiva, triagem e validação de medidas) executados rotineiramente, para garantir que o sistema de

medição esteja em correto funcionamento. Esse processo deve gerar documentos quantitativos sobre a qualidade dos parâmetros medidos (THE METEOROLOGICAL RESOURCE CENTER, 2002). Em relação a instrumentação solarimétrica (piranômetros, pireliômetros e radiômetros) é necessário, a princípio, escolher corretamente o local de instalação e, regularmente, realizar uma inspeção para assegurar que não há objetos projetando sombra sobre o sensor, sujeira e gotas d'água em sua cúpula ou que tenha ocorrido o rompimento do cabeamento.

Antes da aplicação de qualquer procedimento de controle de qualidade das medidas e após a definição de uma rotina regular de manutenção da instrumentação, é necessário conhecer os fatores que afetam os valores da irradiância solar incidente. Younes, Claywell e Muneer (2005) e Muneer, Younes e Munawwar (2007) classificam em duas grandes categorias esses fatores (Figura 7): (a) incertezas e falhas na instrumentação de medida e (b) adversidades operacionais. De forma que as limitações da instrumentação, devido a aspectos de sua construção e desgastes naturais, são as principais fontes de erros nas leituras (WMO, 2012), pois as outras adversidades, de origem operacional, podem ser resolvidas por meio de manutenção preventiva e de treinamento de recursos humanos.

Figura 7 - Fatores que afetam as medidas da irradiância solar.

INCERTEZAS NA INSTRUMENTAÇÃO	ADVERSIDADES OPERACIONAIS
 <p>RESPOSTA DO COSSENO É a resposta do sensor ao ângulo em que a radiação solar atinge a área de detecção.</p>	 <p>POEIRA E SUJEIRA Poeira, sujeira, neve ou qualquer outro material sob a cúpula do sensor pode bloquear os raios solares incidentes.</p>
 <p>RESPOSTA DO AZIMUTE É o resultado de imperfeições nas cúpulas dos instrumentos de radiação solar.</p>	 <p>ORVALHO E GOTAS D'ÁGUA A umidade do ar condensada ou gotas d'água na cúpula do sensor altera o valor de radiação solar captado.</p>
 <p>RESPOSTA DA TEMPERATURA É uma falha nos componentes do sensor causada por grandes flutuações na temperatura.</p>	 <p>EXCREMENTO DE AVES O excremento (fezes ou urina) animal, principalmente de aves, pode bloquear o campo de captação do sensor.</p>
 <p>SELETIVIDADE ESPECTRAL É uma falha causada pela dependência espectral da tinta preta do sensor e da transmissão espectral da cúpula.</p>	 <p>NIVELAMENTO DO SENSOR O incorreto nivelamento da instrumentação pode gerar valores não condizentes com a incidência solar na superfície desejada.</p>
 <p>ESTABILIDADE DO APARELHO É uma falha causada pela deterioração do material do sensor, que gera uma perda de sensibilidade.</p>	 <p>SOMBRA DE ESTRUTURAS Sombras projetadas por estruturas no local de medição da radiação solar podem gerar valores incorretos.</p>
 <p>NÃO-LINEARIDADE É o desvio de sensibilidade da baixa intensidade (100 W/m^2) para a de alta intensidade (1000 W/m^2) no sensor.</p>	 <p>ROMPIMENTO DOS CABOS Ruído ou rompimento do cabeamento de ligação dos sensores podem gerar valores anômalos na série temporal.</p>
 <p>ANEL SOMBREADOR É uma falha causada pelo desalinhamento do anel sombreador sob o sensor.</p>	 <p>ELETRICIDADE Uma falha no fornecimento de energia elétrica pode causar o desligamento do datalogger e de outros dispositivos.</p>

Fonte: Adaptado de Younes, Claywell e Muneer (2005) e Muneer, Younes e Munawwar (2007).

A triagem e validação de valores medidos de irradiância solar são procedimentos que devem ser realizados por meio da aplicação de testes sequenciais a partir de algoritmos de computador. Esses procedimentos têm como objetivo principal filtrar e validar as medições obtidas a partir de diferentes princípios. Os mais

comuns tem como intuito (OHMURA et al., 1998): verificar se os valores estão dentro dos limites fisicamente possíveis e raros (intervalos de valores), analisar a relação entre os valores medidos por sensores diferentes no mesmo ambiente (diferença percentual entre as fontes distintas), comparar as medidas com valores obtidos por meio de modelos (desvio relativo entre os dados) e acompanhar o comportamento sazonal (de acordo com as estações do ano).

O Laboratório Nacional de Energias Renováveis (NREL) do Departamento de Energia dos Estados Unidos da América, em 1993, foi uma das primeiras organizações a promover e divulgar um conjunto de recomendações de controle de qualidade de irradiâncias solares. Seu objetivo era identificar medições fisicamente impossíveis, detectando apenas valores excessivamente errôneos (NREL, 1993). Posteriormente, a Comissão Internacional de Iluminação (CIE), instituição normativa internacional, em 1994, publicou um relatório técnico promovendo um novo procedimento automatizado, composto por etapas que vão desde a delimitação do ângulo mínimo de incidência solar até a comparação de valores com limites físicos e medidas realizadas por outros sensores no local (TREGENZA et al., 1994).

Esses procedimentos nortearam durante anos o controle de qualidade de sistemas de medição da irradiância solar na superfície terrestre. Porém continham uma limitação evidente, eles identificavam apenas erros que causavam grandes variações e ignoravam oscilações errôneas menores nos valores obtidos. No intuito de superar essa limitação, diferentes agências meteorológicas criaram métodos complementares, o que gerou novas perspectivas na detecção de adversidades. Como é o caso do Centro Mundial de Monitoramento de Radiação (WMRC) que produziu um procedimento com métodos para garantir a qualidade das medições da Rede de Radiação em Superfície Terrestre (BSRN), e se consolidou como uma referência em procedimentos de controle de qualidade (LONG; DUTTON, 2002).

Baseado na Rede de Radiação em Superfície Terrestre (BSRN), o Instituto Meteorológico Real da Bélgica (RMIB) criou um procedimento de controle de qualidade de suas medidas da irradiância solar com foco na razão física (detecção de eventos fisicamente impossíveis) e na variabilidade estatística (detecção de eventos raros) dos valores (JOURNÉE; BERTRAND, 2011). É importante destacar que, apesar de todos os procedimentos citados avaliarem a qualidade das medições da irradiância solar, não é recomendado apagar os valores detectados como errôneos da série de medidas. Portanto, ambos são implementados para somente sinalizar quando uma

medida é reprovada em uma das regras de checagem, tal que incumbe ao responsável pelo processo a decisão final de utilizar ou não uma leitura demarcada.

2.5 Preenchimento de Lacunas de Radiação Solar

Uma lacuna é um espaço vazio em uma série de medidas, geralmente, causado por uma falha no sistema de medição. As falhas mais comuns em sistemas terrestres de medição da radiação solar são causadas por excremento de aves encobrendo o sensor, sombras projetadas por estruturas próximas da instrumentação, rompimentos dos cabos e alimentação elétrica dos aparelhos. Isto é, situações do dia a dia que impactam na aquisição de uma medida. No entanto, as lacunas também podem ser causadas por uma medida de má qualidade que foi removida da série temporal após a aplicação de um procedimento de controle de qualidade. Independente do motivo, se um valor não foi coletado, ou foi removido, preencher a lacuna deixada por ele é importante para se ter séries de medições completas.

Existem, na literatura, diferentes metodologias para se estimar valores e usá-los para preencher lacunas em séries de medidas da irradiância solar e de outros parâmetros climáticos, no geral, visando corrigir as falhas do sistema de medição. Baseando-se no estudo de Voyant et al. (2017), é possível segmentar em quatro categorias as metodologias que podem ser usadas para o preenchimento de lacunas em séries temporais climáticas:

- a)** Modelagem baseada na razão física.
- b)** Modelagem baseada em técnicas estatísticas univariadas e multivariadas.
- c)** Modelagem obtida por sensoriamento remoto.
- d)** Modelagem baseada em inteligência artificial.

A modelagem física é uma metodologia baseada na relação física que existe entre as componentes da irradiância solar que se deseja estimar. Já a modelagem estatística compreende técnicas univariadas e multivariadas, aplicadas em função de parâmetros climáticos e geográficos da irradiância solar que se deseja obter valores. Tal que, a modelagem feita por sensoriamento remoto é baseada em produtos oriundos de satélites e de outras informações coletadas por sistemas remotos de medição de parâmetros físicos. Assim como, a modelagem baseada em inteligência artificial compreende técnicas estatísticas aprimoradas por meio de recursos computacionais. A decisão sobre qual metodologia utilizar deve ser tomada de acordo com a quantidade de medidas que se possui e das ferramentas disponíveis.

Com relação ao emprego de metodologias baseadas em estatística univariada e multivariada, Escobedo et al. (2011) usaram a técnica de regressão linear simples para se estimar valores diários das irradiâncias solares ultravioleta, visível e infravermelha. Usando a técnica de análise de componentes principais, Zarzo & Martí (2011) também estimaram valores diários, porém somente da irradiação solar global para o preenchimento de lacunas. Já na área de climatologia, Ventura et al. (2013) utilizaram as técnicas de média aritmética, média móvel, regressão linear simples e regressão linear múltipla para se estimar valores meteorológicos sub-horários. Por fim, Machado e Rampinelli (2018), exploraram as técnicas probabilísticas de replicação e persistência para se estimar valor sub-horários da irradiância solar global.

Levando em consideração as metodologias baseadas na razão física, assim como em metodologias com ênfase em sensoriamento remoto, Janjai, Wattan e Sripradit (2015), estimaram a irradiância solar visível a partir de modelos fundamentados por parâmetros físicos obtidos por satélites. Seguindo a mesma linha de raciocínio, Charuchitipan et al. (2018) criaram um modelo de estimativa da irradiância solar infravermelha, com base em parâmetros físicos obtidos por satélites em correlação com medições obtidas na superfície terrestre. Por fim, Qin et al. (2020), em um estudo mais recente, criaram uma versão melhorada do modelo clássico de Leckner (1978), com base em parâmetros físicos para a estimativa de valores da componente ultravioleta das irradiâncias solares espectrais.

Tendo em vista as metodologias baseadas em inteligência artificial, Nwokolo, Ogbulezie e Obwulu (2022) criaram modelos a partir de cinco algoritmos de Machine Learning (Bagging, Boosting, Rede Neural Artificial, ARIMA e ARIMA Controlada) para se estimar a irradiância solar visível. Seguindo uma proposta híbrida, utilizando uma técnica mais aprimorada, Gallo et al. (2022) criaram dois modelos baseados em algoritmos de Deep Learning (3D-CNN e LSTM) para extrair informações de imagens de satélite e estimar a irradiância solar global. Por fim, em uma compilação do estado da arte, Voyant et al. (2017) reuniram uma série de estudos que usam modelos de Machine Learning para prever a irradiância solar global, baseados em algoritmos de aprendizado supervisionado, não-supervisionado e por reforço.

Um algoritmo de Machine Learning de aprendizado supervisionado se caracteriza por utilizar dados rotulados, associados a uma categoria ou a um valor de interesse, de forma que compreende a relação entre eles e a saída que deve gerar, a fim de fazer previsões. O aprendizado não-supervisionado tem como característica

principal um algoritmo que usa dados não rotulados, de modo que, sem supervisão, consiga organizá-los e descrevê-los de alguma forma, com base em semelhanças entre eles. O aprendizado por reforço é caracterizado por um algoritmo que escolhe uma ação em resposta a uma situação, de forma que recebe um sinal de recompensa, que indica que a decisão foi boa ou ruim, para, então, criar uma estratégia para alcançar, sempre, a recompensa mais alta (MICROSOFT LEARN, 2022a).

O processo de criação de um modelo de Machine Learning, baseado em aprendizado supervisionado, usando uma plataforma de computação em nuvem, é composto por quatro etapas, de acordo com a Microsoft Azure (2022b):

- a) Preparação dos Dados:** é a etapa onde deve-se compilar, identificar e tratar anomalias na série de dados, assim como desenvolver a estrutura que será usada para se treinar o modelo de Machine Learning.
- b) Treinamento do Modelo:** é a etapa onde os dados são segmentados em duas parcelas, onde uma é usada para treinar e a outra para validar o modelo. Nesta etapa deve-se definir os algoritmos que serão usados.
- c) Validação do Modelo:** é a etapa onde uma parte dos dados deverá ser utilizada para avaliar o desempenho do modelo que foi treinado, a partir das métricas escolhidas.
- d) Implantação do Modelo:** é a etapa onde deve-se escolher o modelo que gerou os melhores resultados a partir das métricas definidas e implantá-lo como um serviço disponível na internet para fazer previsões (API Rest).

3 MATERIAL E MÉTODOS

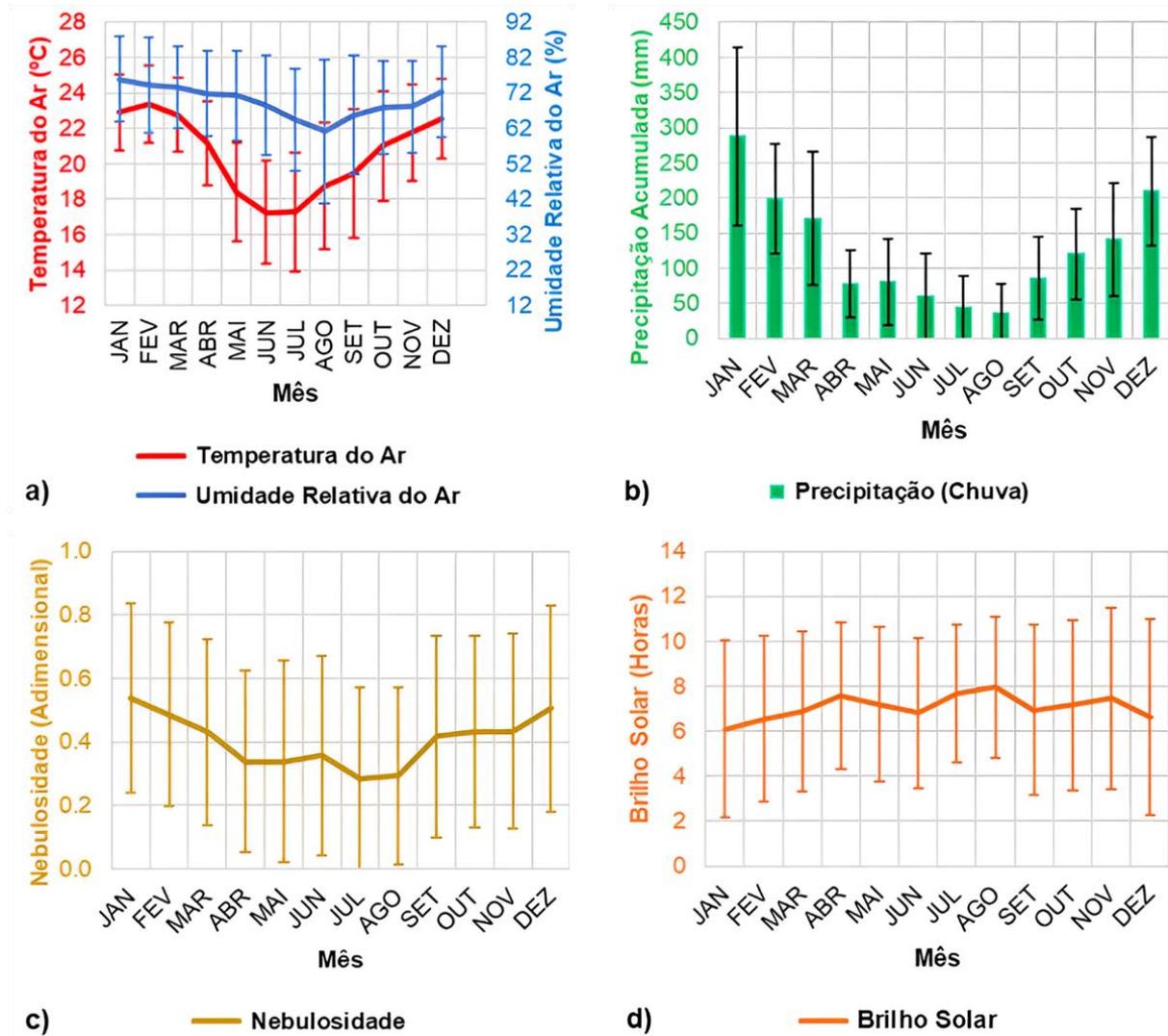
Este estudo foi conduzido utilizando medições experimentais (Seção 3.2) do Laboratório de Agrometeorologia e Radiometria Solar (22°54'S de latitude, 48°27'O de longitude e 786m de altitude) da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP de Botucatu (SP) - Brasil (Seção 3.1). O intuito da proposta foi criar um procedimento de detecção de medidas errôneas e isoladas, por meio do controle de qualidade, das irradiâncias solares ultravioleta, visível e infravermelha em função dos princípios físicos (Seção 3.3.1) e de intervalos estatísticos (Seção 3.3.2), assim como, elaborar modelos de estimativa e preenchimento de lacunas das componentes espectrais (Seção 3.4) em resolução temporal sub-horária (5 minutos em W/m^2).

3.1 Localização e Clima

Botucatu é um município situado na região centro-sul do estado de São Paulo, que possui clima temperado, caracterizado por um inverno seco (de junho a agosto) e por um verão quente e úmido (de dezembro a fevereiro), com fortes atividades agrícolas, que compreendem grandes plantações de cana-de-açúcar e eucalipto, bem como atividades industriais moderadas (CODATO et al., 2008; DAL PAI et al., 2016). Sua formação geográfica compreende um relevo de colinas não simétricas, conhecido como Cuesta de Botucatu, caracterizado por dois níveis distintos de altitudes, na faixa de 500 a 400 metros na região mais baixa, e de 700 a 900 metros na região serrana, mais alta. Esta formação provoca diferenças na temperatura do município (SILVA et al., 2017).

A série histórica de medidas climáticas para o município, obtidas pela Estação Meteorológica Lageado (2022), aponta o mês de fevereiro como o mais quente, com temperatura média do ar de 22,5 °C, e julho como o mais frio, com média de 16,8 °C (Figura 8a). Janeiro é o mês com maior acúmulo pluviométrico, com média de 300 mm, e agosto com o menor, com média de 50 mm (Figura 8b). Sendo assim, janeiro é o mês mais úmido, com 75,6% de umidade relativa média do ar, e agosto o menos, com 61,20% (Figura 8a), também na média. Conseqüentemente, janeiro é o mês mais nebuloso, com um índice médio de 0,54 de cobertura do céu, e julho o menos, com um índice médio de 0,28 de cobertura do céu (Figura 8c). Tais informações consideram os valores médios aferidos entre os anos de 1971 e 2013.

Figura 8 - Série histórica de medidas climáticas da Estação Meteorológica Lageado (2022) .



Fonte: Estação Meteorológica Lageado (2022).

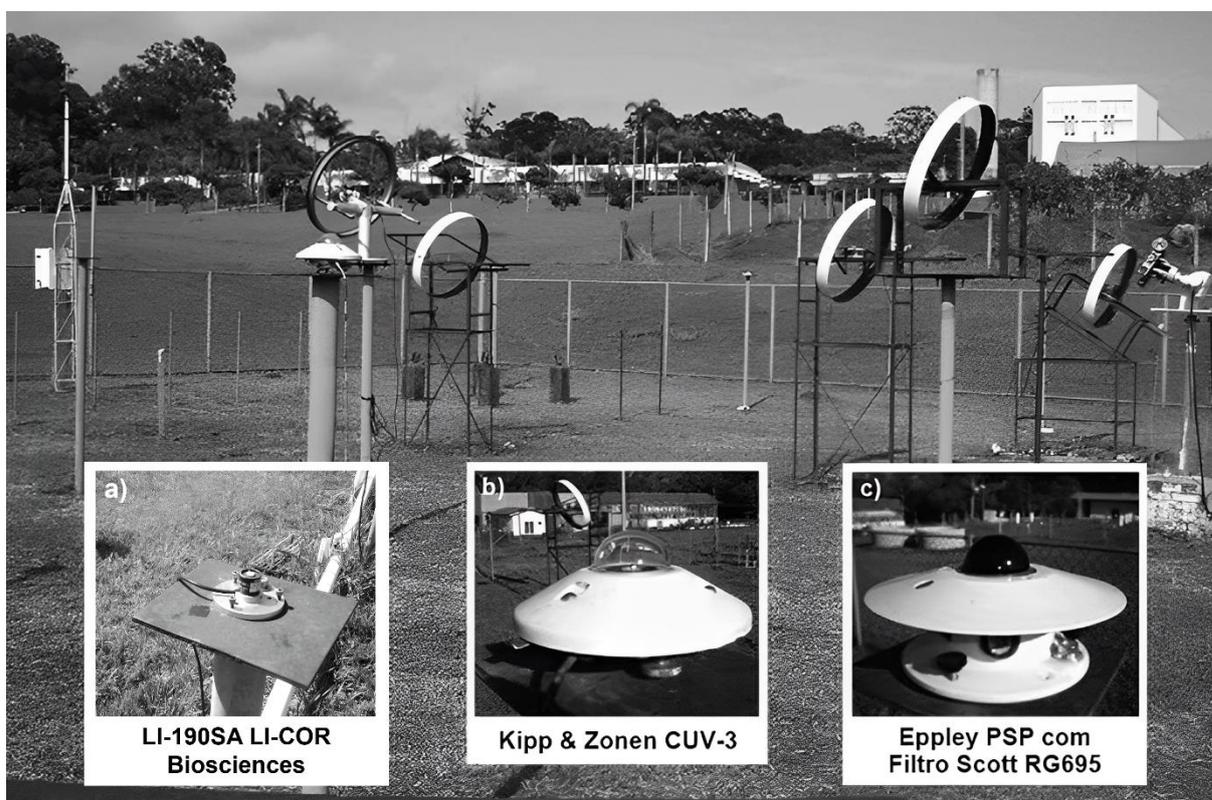
A nebulosidade é um índice $(1 - n/N)$ que representa a quantidade de horas que o Sol ficou encoberto durante o dia. Segundo Iqbal (1983), ela representa a razão de insolação, que pode ser obtida a partir das horas de brilho solar (n), com valor mínimo de 6,1 horas em janeiro e máximo de 7,9 horas em agosto (Figura 8d) no município de Botucatu (SP), e do fotoperíodo (N).

3.2 Instrumentação e Medidas

Foi utilizada a série temporal completa (2001 a 2006) de medidas diurnas em superfície terrestre horizontal das irradiâncias solares ultravioleta (UV), mensurada por um radiômetro CUV-3 da Kipp & Zonen (Figura 9b), infravermelha-próxima (IV), mensurada por um piranômetro Eppley PSP com cúpula de transmissão seletiva

(Filtro Scott RG695) na faixa espectral de 700 a 3000 nm (Figura 9c), e visível, também conhecida como fotossinteticamente ativa (PAR), obtida por meio de um sensor quântico LI-190SA da LI-COR Biosciences (Figura 9a). Após a realização do experimento de obtenção das componentes espectrais da irradiância solar, no período em questão, Escobedo et al. (2006), Escobedo et al. (2009), Escobedo et al. (2011) e Corrêa (2011), utilizaram essas medidas em estudos científicos de modelagem.

Figura 9 - Radiômetros do Laboratório de Agrometeorologia e Radiometria Solar.



Fonte: Adaptado de Escobedo et. al (2011).

Segundo Escobedo et al., (2011) a calibração dos instrumentos utilizados durante o período de medição em questão foi realizada a cada dois anos de acordo com as recomendações da World Meteorological Organization (WMO). Assim como, seguindo as recomendações do fabricante, o fator de calibração do radiômetro CUV-3 da Kipp & Zonen (2004) foi multiplicado por 0,92 para corrigir possíveis influências do filtro de transmissão sob seu elemento sensível. Na Tabela 1 estão descritas as propriedades técnicas de operação dos instrumentos, referentes ao fator de calibração de fábrica, intervalo espectral, tempo de resposta, linearidade, resposta ao cosseno e a temperatura do elemento sensível.

Tabela 1 - Propriedades operacionais dos instrumentos de medição das irradiâncias solares espectrais.

INSTRUMENTO	IRRADIÂNCIA SOLAR		
	VISÍVEL	ULTRAVIOLETA	INFRAVERMELHA
Fabricante (Modelo)	Eppley (PSP)	Kipp & Zonen (CUV-3)	Eppley (PSP)
Fator de Calibração	5,29 $\frac{\mu A}{1000} \mu mol s^{-1} m^{-2}$	312 $\mu V/Wm^2$	8,12 $\mu V/Wm^2$
Intervalo Espectral	400 a 700nm	290 a 400nm	695 a 2800nm
Tempo de Resposta	10 μs	5ms	1s
Linearidade	$\pm 1\%$ acima de 1000 $\mu mol s^{-1} m^{-2}$	< 1%	$\pm 1\%$ de 0 a 700 W/m^2
Resposta do Cosseno	Acima de 80°	$\pm 10\%$	$\pm 5\%$
Resposta a Temperatura	$\pm 0,15\%$ por °C	$\pm 0,1/K$	$\pm 1\%$ de - 20°C a 40°C

Fonte: Kipp & Zonen (2004); Escobedo et al. (2011), Eppley (2011) e LI-COR Biosciences (2022).

Durante o período de realização das medições experimentais foi utilizado um data logger modelo CR23X Micrologger da Campbell Scientific (2006), configurado para escanear as medidas das irradiâncias solares a cada 5 segundos e armazenar seus valores médios a cada 5 minutos, o que gerava 288 valores mensurados durante o intervalo de um dia. Diariamente as medidas coletadas foram gravadas em um módulo externo de memória modelo SM192 da Campbell Scientific (1993) e semanalmente

eram transferidas para um servidor de dados por meio de uma interface SC532 e do software PC208W da Campbell Scientific (2001).

3.3 Controle de Qualidade

Investigando a literatura científica que envolve estudos sobre a utilização de medidas das irradiâncias solares espectrais, foi possível constatar que não existe um método de controle de qualidade convencionalmente aceito para a detecção de valores errôneos ou isolados das irradiâncias solares ultravioleta, visível e infravermelha. No entanto, Corrêa (2011) cita que para filtrar ruídos, valores ausentes e noturnos, em suas medidas, usou um sistema computacional, sem detalhar minuciosamente os procedimentos adotados. Assim como, Escobedo et al. (2009) e Escobedo et al. (2011), detalham que eliminaram medidas obtidas no começo e no final de todos os dias da série temporal, devido aos efeitos de sombreamento nos sensores. Além destas informações nenhum processo mais técnico foi divulgado.

Entende-se, neste caso, que o controle de qualidade é um processo composto por procedimentos de calibração de instrumentos, inspeções físicas, manutenção preventiva, detecção de medidas inconsistentes e análises estatísticas, que garantem que o sistema de medição esteja em correto funcionamento (OHMURA et al., 1998; CALÇA et al., 2019). Tal que, o foco deste estudo é propor um novo procedimento de controle de qualidade para a detecção de medidas sub-horárias errôneas ou isoladas das irradiâncias solares ultravioleta, visível e infravermelha, com base em razões físicas e análise estatística. O intuito não é apagar informações da série de medidas, mas sinalizar os valores errôneos ou isolados dos demais com um código alfabético (Tabela 2), que informará se ele é válido (correto) ou inválido (incorreto).

Tabela 2 - Diagnósticos do procedimento de controle de qualidade de acordo com o código numérico atribuído ao valor medido.

CÓDIGO ALFABÉTICO	DIAGNÓSTICO
INV	Valor inválido (incorreto)
VAL	Valor válido (correto)

A sinalização dos valores errôneos ou isolados na série de medidas, permite que se possa analisar quantitativamente a natureza das adversidades, e que se decida usá-los ou não em outros estudos. Neste sentido, o procedimento foi

implementado a partir de um algoritmo computacional codificado na linguagem de consulta estruturada (SQL) por meio do sistema de gerenciamento de banco de dados relacional (SGBDR) MySQL. O controle de qualidade se baseou em três princípios físicos, para identificar valores errôneos, adaptados da Baseline Surface Radiation Network (BSRN) de Long e Dutton (2002), e em uma análise estatística univariada, para identificar valores isolados (outliers), em função do quantil da distribuição normal que estabelece a proporção de valores dispostos em torno da média.

3.3.1 Princípios Físicos de Detecção de Medidas Errôneas

Em virtude das incertezas no processo de obtenção das irradiâncias solares espectrais, se torna necessária a utilização de verificações baseadas no raciocínio físico, a fim de identificar valores impossíveis, ou seja, medidas consideradas errôneas (JOURNÉE; BERTRAND, 2011). Portanto, a partir das equações 1, 2 e 3 foram verificadas as medidas que se encontravam dentro dos limites fisicamente possíveis (primeiro princípio físico), bem como com as equações 4, 5 e 6 foram analisados os valores que estavam dentro dos limites considerados raros (segundo princípio físico) e, por fim, por meio das equações 7, 8 e 9 foram feitas comparações com os valores obtidos para o topo da atmosfera (terceiro princípio físico) para as irradiâncias solares ultravioleta, visível e infravermelha, respectivamente.

a) Verificação dos limites fisicamente possíveis (**Primeiro Princípio**):

$$-4 \text{ W/m}^2 \leq UV \leq I_{sc} E_{xc} 1,5 \text{ Cos}(Z)^{1,2} + 100 \text{ W/m}^2 \quad (1)$$

$$-4 \text{ W/m}^2 \leq PAR \leq I_{sc} E_{xc} 1,5 \text{ Cos}(Z)^{1,2} + 100 \text{ W/m}^2 \quad (2)$$

$$-4 \text{ W/m}^2 \leq IV \leq I_{sc} E_{xc} 1,5 \text{ Cos}(Z)^{1,2} + 100 \text{ W/m}^2 \quad (3)$$

b) Verificação dos limites extremamente raros (**Segundo Princípio**):

$$-2 \text{ W/m}^2 \leq UV \leq I_{sc} E_{xc} 1,2 \text{ Cos}(Z)^{1,2} + 50 \text{ W/m}^2 \quad (4)$$

$$-2 \text{ W/m}^2 \leq PAR \leq I_{sc} E_{xc} 1,2 \text{ Cos}(Z)^{1,2} + 50 \text{ W/m}^2 \quad (5)$$

$$-2 \text{ W/m}^2 \leq IV \leq I_{sc} E_{xc} 1,2 \text{ Cos}(Z)^{1,2} + 50 \text{ W/m}^2 \quad (6)$$

c) Comparação com os valores do topo da atmosfera (**Terceiro Princípio**):

$$0 \text{ W/m}^2 < UV \leq (UV_0 1.20) \quad (7)$$

$$0 \text{ W/m}^2 < PAR \leq (PAR_0 1.20) \quad (8)$$

$$0 \text{ W/m}^2 < IV \leq (IV_0 \cdot 1.20) \quad (9)$$

Nas equações de 1 a 9 o termo UV, PAR e IV representam as irradiâncias solares ultravioleta, visível e infravermelha (W/m^2), respectivamente, I_{sc} é a constante solar (W/m^2) calculada para as irradiâncias solares (Tabela 3), E_{xc} (Equação 10) é o fator de correção da excentricidade da órbita terrestre (adimensional), Z (Equação 11) é o ângulo zenital (graus), UV_0 , PAR_0 e IV_0 são as irradiância solares ultravioleta, visível e infravermelha no topo da atmosfera (W/m^2), respectivamente. As medições das componentes espectrais da irradiância solar foram analisadas sequencialmente, quando seus valores não estavam condizentes com o limite determinado por algum princípio físico eram sinalizados como inválidos, não sendo submetidos a demais análises subsequentes, quando seus valores estavam de acordo com o limite determinado por algum dos princípios físicos eram sinalizados como válidos, permitindo que fossem verificados pelas análises subsequentes.

Tabela 3 - Constante solar das irradiâncias solares global, ultravioleta, visível e infravermelha.

IRRADIÂNCIA SOLAR	CONSTANTE SOLAR (I_{sc})
Global	1361 W/m^2
Ultravioleta	77,577 W/m^2
Visível	528,068 W/m^2
Infravermelha	755,355 W/m^2

As constantes solares das irradiâncias solares ultravioleta, visível e infravermelha foram calculadas com base no percentual que cada uma representa da irradiância solar global. De forma que a ultravioleta representa 5,7%, a visível representa 38,8% e a infravermelha representa 55,5% do espectro da irradiância solar global, de acordo com Corrêa (2011). Dessa forma pôde-se obter também os valores das irradiâncias solares ultravioleta, visível e infravermelha para o topo da atmosfera, a partir da equação 14, utilizada para calcular a irradiância solar global no topo da atmosfera (G_0), apenas alterando a sua constante solar.

$$E_{xc} = 1 - 0,0009467 \text{ sen } (F) - 0,01671 \text{ cos } (F) - 0,0001489 (2 F) - 0,00002917 \quad (10)$$

$$\text{sen } (3 F) - 0,0003438 \cos (4 F)$$

$$Z = \text{sen } (\delta) \text{sen } (\phi) + \cos (\delta) \cos (\phi) \cos (\omega) \quad (11)$$

Nas equações 10 e 11 o termo F representa $360 D/365$, onde D é o dia decorrido do ano do 1º ao 365º, δ é a declinação solar em graus (Equação 12), ϕ é a latitude geográfica do local onde foram obtidas as irradiâncias solares na Fazenda Lageado - Botucatu (SP) em graus e ω é o ângulo horário em graus (Equação 13).

$$\delta = 0,3964 + 3,631 \text{sen } (F) - 22,97 \cos (F) + 0,03838 \text{sen } (2 F) - 0,3885 \cos (2 F) + 0,07659 \text{sen } (3 F) - 0,1587 \cos (3 F) - 0,01021 \cos (4 F) \quad (12)$$

$$\omega = (12 - H_d) 15 \quad (13)$$

$$G_0 = I_{sc} \text{Exc} \cos (Z) \quad (14)$$

Na equação 13 o termo H_d representa a hora e o décimo de hora do dia no formato decimal. Todas as equações para obtenção das efemérides solares/geoespaciais utilizadas neste estudo foram obtidas de acordo com os cálculos propostos por Iqbal (1983) em sua obra mais clássica da área de radiação solar.

3.3.2 Intervalo Estatístico de Detecção de Medidas Isoladas

Em virtude das incertezas no processo de medição das irradiâncias solares espectrais, é necessário também a utilização de procedimentos baseados em técnicas estatísticas, a fim de identificar eventos questionáveis na série de medidas, ou seja, a ocorrência de valores isolados dos demais (JOURNEÉ; BERTRAND, 2011). Popularmente conhecidos como outliers, valores isolados são aqueles que se desviam consideravelmente das demais observações de uma série temporal, causando suspeitas quanto a sua veracidade (WANG, 2019). Em vista disso, no sentido de gerar uma análise complementar aos princípios físicos de identificação de medidas errôneas, foi realizada a construção de intervalos estatísticos de confiança.

Os intervalos estatísticos de confiança devem determinar as faixas de variações admissíveis para as medições da série temporal, de forma que seus limites extremos demarcam os valores máximos e mínimos que uma observação deve atingir, atestando sua confiabilidade. Considerando que a parcela das irradiâncias solares ultravioleta, visível e infravermelha que chega na superfície terrestre depende das condições do ambiente, foi feita uma classificação da cobertura de céu (Tabela 4) por meio do índice de transmissividade atmosférica (KT), que representa a razão entre a irradiância solar global medida e a extraterrestre calculada (IG/IE). Para que, desta forma, fosse possível gerar intervalos de confiança considerando a condição que mais afeta a transmissão dos raios solares, que é a nebulosidade.

Tabela 4 - Classificação da cobertura de céu em função do índice de transmissividade atmosférica.

CLASSIFICAÇÃO DA COBERTURA DE CÉU	TRANSMISSIVIDADE ATMOSFÉRICA
Aberto	$0,65 \leq KT < 1$
Parcialmente Aberto	$0,55 \leq KT < 0,65$
Parcialmente Nublado	$0,35 \leq KT < 0,55$
Nublado	$0 \leq KT < 0,35$

Fonte: Escobedo et al., 2009.

Portanto, para a construção dos intervalos estatísticos de detecção de medidas isoladas foram obtidas informações como a quantidade, a média, o desvio padrão, o valor mínimo e o valor máximo para toda a série de irradiâncias solares espectrais considerando o padrão de cobertura de céu. De modo que, só se considerou medições não sinalizadas como inválidas, que passaram pelas três regras anteriores, relacionadas aos princípios físicos de controle de qualidade. Outro parâmetro fundamental é o nível de confiança (Tabela 5), que indica a probabilidade de um valor estar entre os limites inferiores e superiores de variação de acordo com a distribuição normal de probabilidade, o que sinaliza estar coerente com as demais observações da série de medidas (ALBERTAZZI; SOUSA, 2018; BONAFARI, 2012).

Tabela 5 - Intervalos estatísticos de acordo com o nível de confiança considerando a distribuição normal de probabilidade.

INTERVALO	NÍVEL DE CONFIANÇA
$\mu \pm 1,00\sigma$	68,20%
$\mu \pm 2,00\sigma$	95,40%
$\mu \pm 3,00\sigma$	99,70%

Fonte: Carrilho e Galo (2017).

Procedimentos estatísticos semelhantes, baseados em outros parâmetros climáticos como a transmissividade atmosférica, coeficiente de difusão e fator de turbidez, foram propostos nos estudos de Younes, Claywell e Muneer (2005), Journée e Bertrand (2011) e Pashiardis e Kalogirou (2016), utilizando intervalos de 68,20% ($\mu \pm 1,00\sigma$) a 99,70% ($\mu \pm 3,00\sigma$) de confiança e outras informações estatísticas como a média (μ) e o desvio padrão (σ). Em síntese, neste estudo foi utilizado o valor médio e o de desvio padrão das irradiâncias solares ultravioleta, visível e infravermelha, com base nas estatísticas obtidas para cada classificação da cobertura de céu, além do nível de 99,70% de confiança para a construção dos intervalos estatísticos de cada componente espectral (Equações 15, 16 e 17).

$$\mu_{UV} - (3,00 \sigma_{UV}) \leq UV \leq \mu_{UV} + (3,00 \sigma_{UV}) \quad (15)$$

$$\mu_{PAR} - (3,00 \sigma_{PAR}) \leq PAR \leq \mu_{PAR} + (3,00 \sigma_{PAR}) \quad (16)$$

$$\mu_{IV} - (3,00 \sigma_{IV}) \leq IV \leq \mu_{IV} + (3,00 \sigma_{IV}) \quad (17)$$

Onde μ representa a média e σ o desvio padrão populacional das medições que não foram sinalizadas como errôneas pelos princípios físicos no período de 2001 a 2006 e 3 é o quantil que estabelece o nível de confiança. Neste sentido, foram considerados isolados os valores das irradiância solares espectrais que estavam fora dos limites de variação propostos no intervalo estatístico, considerando um nível de confiança de 99,70% de acordo com a distribuição normal de probabilidade. Pode-se dizer, portanto, que mesmo após os valores medidos terem sido validados pelos princípios físicos de controle de qualidade, ou seja, não terem ainda sido detectados

como errôneos, podem ser invalidados pelo intervalo estatístico, caso não estejam dentro dos valores de variação aceitos.

3.4 Preenchimento de Lacunas com Machine Learning

Uma lacuna é um espaço vazio, causado por uma falha no sistema de medição ou por uma medida de má qualidade que foi removida da série temporal por algum motivo técnico. Neste sentido, quando se trata de monitorar as irradiâncias solares espectrais, o processo de medição é mais laborioso do que o de outros parâmetros climáticos, devido às incertezas causadas pela instrumentação e as adversidades operacionais. Portanto, para se preencher lacunas na série de medidas, podem ser utilizados modelos estruturais baseados em outras variáveis meteorológicas e parâmetros geográficos, modelos de séries temporais baseados em observações históricas de um mesmo conjunto de medições, modelos baseados em observações históricas e em outras variáveis meteorológicas ou modelos de Machine Learning baseados em variáveis explicativas que geram um valor de saída (AGGARWAL; SAINI, 2014; VOYANT et al., 2017).

Portanto, foram criados neste estudo modelos de Machine Learning (inteligência artificial) utilizando os recursos visuais automatizados da plataforma de computação em nuvem Microsoft Azure, por meio do serviço Azure Machine Learning, com base no Microsoft Learn (2022b). A Microsoft Azure é uma plataforma de computação em nuvem comercial que fornece acesso, gerenciamento e utilização de recursos computacionais (máquinas virtuais, desenvolvimento de aplicativos de software, gerenciamento de bancos de dados e criação de serviços especializados) por meio de data centers distribuídos em todo o mundo, onde se tem o pagamento baseado no uso real de determinados recursos computacionais, assim como se tem a oferta de serviços gratuitos. Neste estudo, optou-se por utilizar o acesso de estudante, onde a Microsoft Azure fornece um total de US\$ 100,00 para uso em todos os serviços de computação em nuvem, com renovação anual.

Apesar de ser uma plataforma de computação em nuvem que cobra por uma grande parte dos serviços disponibilizados ao público, a grande vantagem da Microsoft Azure é que durante a etapa de treinamento dos modelos, o serviço Azure Machine Learning cria diferentes processos iterativos usando computação paralela. Esses processos experimentam os parâmetros e os algoritmos existentes na plataforma de computação em nuvem, todos ao mesmo tempo, e produzem resultados

para os diferentes modelos de Machine Learning, exibindo-os a partir de um ranking com as métricas de avaliação. Neste sentido, foram criados modelos para estimar as componentes espectrais a cada 5 minutos (W/m^2) com base em métodos de Machine Learning de regressão na Microsoft Azure, de forma que podem ser usados para preencher lacunas das irradiâncias solares ultravioleta, visível e infravermelha, devido a capacidade que possuem de estimar valores dessas componentes espectrais.

No Azure Machine Learning, de acordo com a Microsoft Learn (2022a), denominam-se recursos as variáveis independentes, cujo valor influencia aquele que se quer prever e rótulo de previsão a variável dependente, que é exatamente o que se deseja prever. Neste sentido, para se criar um modelo de Machine Learning que produza bons resultados, é necessário que todas as linhas do conjunto de dados de treinamento possuam algum valor, e que esse ele seja o mais confiável possível. Por esta razão, foi necessário submeter as medições das componentes espectrais da irradiância solar ao procedimento de controle de qualidade proposto neste estudo (Seção 3.3), pois atestou a confiabilidade das medições utilizadas. Sendo assim, para se criar os modelos de Machine Learning de estimativa das irradiâncias solares ultravioleta, visível e infravermelha, todas as linhas que continham pelo menos um valor errôneo ou isolado foram eliminadas do conjunto de dados de treinamento.

Para treinar o modelo de Machine Learning para estimativa da irradiância solar ultravioleta, do total coletado durante o período de 2001 a 2006 (Seção 3.2), foram utilizadas 293.486 linhas da série de medidas considerando 9 variáveis como recursos e o próprio valor medido da componente espectral como rótulo (Tabela 6a). Em sequência, para o mesmo período foram utilizadas 304.071 linhas da série de medidas para gerar o modelo de Machine Learning para estimativa da irradiância solar visível, também considerando 9 variáveis como recursos e os valores medidos da componente espectral como rótulo (Tabela 6b). Por fim, do total obtido durante o período em questão foram utilizadas 296.860 linhas da série de medidas para gerar o modelo de Machine Learning da irradiância solar infravermelha, considerando novamente, como nos casos anteriores, 9 variáveis como recursos e o valor da componente espectral como rótulo (Tabela 6c).

Tabela 6 - Recursos e rótulo de previsão dos modelos de Machine Learning das irradiâncias solares ultravioleta, visível e infravermelha.

a)	VARIÁVEIS DE TREINAMENTO DO MODELO DE MACHINE LEARNING DA IRRADIÂNCIA SOLAR ULTRAVIOLETA	
RECURSOS		RÓTULO DE PREVISÃO
COLUNA	PROPRIEDADES	
1 ^a	Ano	Irradiância Solar Ultravioleta
2 ^a	Mês	
3 ^a	Dia	
4 ^a	Dia Juliano	
5 ^a	Hora	
6 ^a	Minuto	
7 ^a	Elevação Solar	
8 ^a	Irradiância Solar Global	
9 ^a	Irradiância Solar Ultravioleta Extraterrestre	
b)	VARIÁVEIS DE TREINAMENTO DO MODELO DE MACHINE LEARNING DA IRRADIÂNCIA SOLAR VISÍVEL	
RECURSOS		RÓTULO DE PREVISÃO
COLUNA	PROPRIEDADES	
1 ^a	Ano	Irradiância Solar Visível
2 ^a	Mês	
3 ^a	Dia	
4 ^a	Dia Juliano	
5 ^a	Hora	
6 ^a	Minuto	
7 ^a	Elevação Solar	
8 ^a	Irradiância Solar Global	
9 ^a	Irradiância Solar Visível Extraterrestre	
c)	VARIÁVEIS DE TREINAMENTO DO MODELO DE MACHINE LEARNING DA IRRADIÂNCIA SOLAR INFRAVERMELHA	
RECURSOS		RÓTULO DE

COLUNA	COLUNA	PREVISÃO
1 ^a	Ano	Irradiância Solar Infravermelha
2 ^a	Mês	
3 ^a	Dia	
4 ^a	Dia Juliano	
5 ^a	Hora	
6 ^a	Minuto	
7 ^a	Elevação Solar	
8 ^a	Irradiância Solar Global	
9 ^a	Irradiância Solar Infravermelha Extraterrestre	

De todas as variáveis selecionadas para compor os rótulos de treinamento dos modelos de Machine Learning das componentes espectrais das irradiâncias solares (Tabela 6), as colunas de ano, mês, dia, dia juliano, hora e minuto, representam o tempo, neste sentido, foram incluídas na lista para que posteriormente, pudessem ser usadas como parâmetro de entrada no modelo de estimativa. As colunas de elevação solar, irradiância solar global e irradiância solar extraterrestre (ultravioleta, visível e infravermelha) foram utilizadas pois seus valores influenciam o valor de saída (rótulo de previsão). Isto ocorre, principalmente, com o valor da irradiância solar global, uma vez que as três componentes espectrais utilizadas neste estudo estão relacionadas a quantidade total de irradiância solar que atinge o planeta Terra, ou seja, o total que se tem disponível para aproveitamento na superfície terrestre.

Conforme mencionado, o Azure Machine Learning, permitiu executar, a partir de processamento computacional paralelo, o treinamento de diversos modelos com os algoritmos disponíveis na Microsoft Azure, gerando no final, um ranking com o resultado de cada um em ordem descendente (do melhor para o pior coeficiente de determinação). Seguindo esta lógica, neste estudo, foram avaliados os seguintes algoritmos de Machine Learning, baseados no framework Scikit-Learn (2022b), a partir de um cluster de máquinas virtuais com 6 nós, compreendendo cada um 4 núcleos virtuais de processamento, 8GB de memória RAM e 32GB de armazenamento SSD:

- **Elastic Net** (Modelo Baseado em Regressão Linear).
- **Gradient Boosting** (Modelo Baseado em Árvore de Decisão).

- **Decision Tree** (Modelo Baseado em Árvore de Decisão).
- **KNN** (Modelo Baseado em Classificação).
- **Lasso Lars** (Modelo Baseado em Regressão Linear).
- **SGD** (Modelo Baseado em Regressão Linear).
- **Random Forest** (Modelo Baseado em Árvore de Decisão).
- **Extreme Random Trees** (Modelo Baseado em Árvore de Decisão).
- **Light GBM** (Modelo Baseado em Gradient Boosting).
- **XG Boost Regressor** (Modelo Baseado em Gradient Boosting).
- **Fast Linear Regressor** (Modelo Baseado em Regressão Linear).
- **Gradient Descent Regressor** (Modelo Baseado em Gradient Regressor).

A capacidade de estimar valores das irradiância solares a partir dos modelos criados foi avaliada pelo Azure Machine Learning utilizando também as métricas do framework Scikit-Learn (2022a), por meio dos indicadores estatísticos Mean Absolute Error (MAE), Root Mean Square Error (RMSE) e Coeficiente de Determinação (R^2). O MAE (Equação 18) quantifica a soma dos erros absolutos dividida pelo tamanho da amostra de pares de dados avaliados. O RMSE (Equação 20) estima o erro absoluto indicado pelas diferenças entre os pares de valores avaliados, dando um maior peso para grandes erros. Os valores relativos de MAE (rMAE), obtido pela Equação 19, e de RMSE (rRMSE), obtido pela Equação 21, fornecem o grau de correspondência percentual proporcional à média da observação. De forma que o R^2 (Equação 22) é uma medida de ajuste de um modelo estatístico linear generalizado aos valores observados. Para os indicadores MAE e RMSE valores próximos de zero são desejáveis, porém para o R^2 é mais interessante um valor próximo de um.

$$\mathbf{MAE} = \frac{\sum_{i=1}^n |y_i - y'_i|}{n} \quad (18)$$

$$\mathbf{rMAE} = \left(\frac{\mathbf{MAE}}{\bar{y}} \right) \times 100 \quad (19)$$

$$\mathbf{RMSE} = \left[\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \right] \quad (20)$$

$$r\text{RMSE} = \left(\frac{\text{RMSE}}{\bar{y}} \right) \times 100 \quad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (22)$$

O parâmetro y' representa os dados de comparação (valores modelados), y representa os dados de referência (valores medidos) e \bar{y} é a média dos valores de referência. A validação cruzada k-fold foi o método definido para avaliar o desempenho do modelo de Machine Learning. Neste sentido, no Azure Machine Learning, os dados foram separados aleatoriamente e passaram por 5 rodadas de treinamento e validação. Portanto, 5 treinamentos usando 70% do conjunto de dados foram realizados e 5 validações usando 30% do conjunto de dados foram feitas, o que gerou as métricas definidas (Equações 18, 19, 20, 21 e 22) para cada rodada. O resultado, neste caso, foi gerado por meio da média das 5 rodadas de validação (MICROSOFT LEARN, 2022b).

4 RESULTADOS E DISCUSSÃO

4.1 Análise Exploratória das Irradiâncias Solares Espectrais

Foi possível constatar, a partir da série de irradiâncias solares obtida durante o período de 2001 a 2006, que as componentes espectrais ultravioleta, visível e infravermelha possuem a mesma quantidade de valores medidos com relação a todos os anos em que o experimento foi conduzido (Tabela 7). Neste caso, para cada ano da série temporal o arquivo DAT gerado pelo data logger (Seção 3.2) foi importado para o SGBDR MySQL, desconsiderando linhas duplicadas, para que fosse possível implementar o procedimento de detecção de medidas errôneas e isoladas. Portanto, os valores totais que compreendem a série de medidas das irradiâncias solares espectrais representam seu estado bruto, sendo a contabilização dos registros de qualquer valor obtido, seja ele bom, ruim ou nulo.

Tabela 7 - Quantidade total de medidas das irradiâncias solares ultravioleta, visível e infravermelha sem validação em resolução temporal sub-horária.

ANO	IRRADIÂNCIA SOLAR		
	ULTRAVIOLETA	VISÍVEL	INFRAVERMELHA
2001	51.854	51.854	51.854
2002	51.613	51.613	51.613
2003	51.939	51.939	51.939
2004	51.673	51.673	51.673
2005	51.837	51.837	51.837
2006	52.441	52.441	52.441
TOTAL	311.357	311.357	311.357

Conforme definido neste estudo, uma lacuna é um espaço vazio provocado por uma falha instrumental ao obter uma medida ou pela remoção de um valor errôneo da série temporal. Portanto, é necessário quantificar as lacunas inicialmente causadas por falhas no processo de medição (Tabela 8), pois neste estudo valores nulos na série temporal em estado bruto, não foram contabilizados pelo procedimento de controle de qualidade. Neste sentido, para a componente espectral ultravioleta, a maior quantidade de lacunas foi registrada em 2001 e a menor em 2005. Tal que, para a componente espectral visível a maior quantidade de lacunas foi registrada em 2004

e a menor em 2006. Por fim, para a componente espectral infravermelha, a maior quantidade de lacunas foi registrada em 2004 e a menor em 2003.

Tabela 8 - Quantidade de campos nulos na série de medidas das irradiâncias solares ultravioleta, visível e infravermelha em resolução temporal sub-horária.

ANO	IRRADIÂNCIA SOLAR		
	ULTRAVIOLETA	VISÍVEL	INFRAVERMELHA
2001	9.553	150	2.627
2002	1.194	803	2.168
2003	78	66	1.317
2004	33	1.083	3.029
2005	28	986	1.468
2006	2.231	28	1.381
TOTAL	13.117	3.116	10.743

Após desconsiderar as lacunas registradas, a irradiância solar ultravioleta apresentou em 2003 a maior quantidade de medidas e em 2001 a menor quantidade. Assim como, a irradiância solar visível apresentou em 2006 a maior quantidade de medições e em 2004 a menor quantidade. Por fim, a irradiância solar infravermelha apresentou em 2006 a maior quantidade de valores medidos e em 2004 a menor quantidade. Considerando todos os valores disponíveis para validação durante o período de obtenção de medidas das componentes espectrais da irradiância solar (Tabela 9) a visível é a que tem a maior e a ultravioleta a menor quantidade de valores registrados, respectivamente. Neste caso, compreende-se, que todos os valores representam medidas em que sua veracidade precisa ser inspecionada.

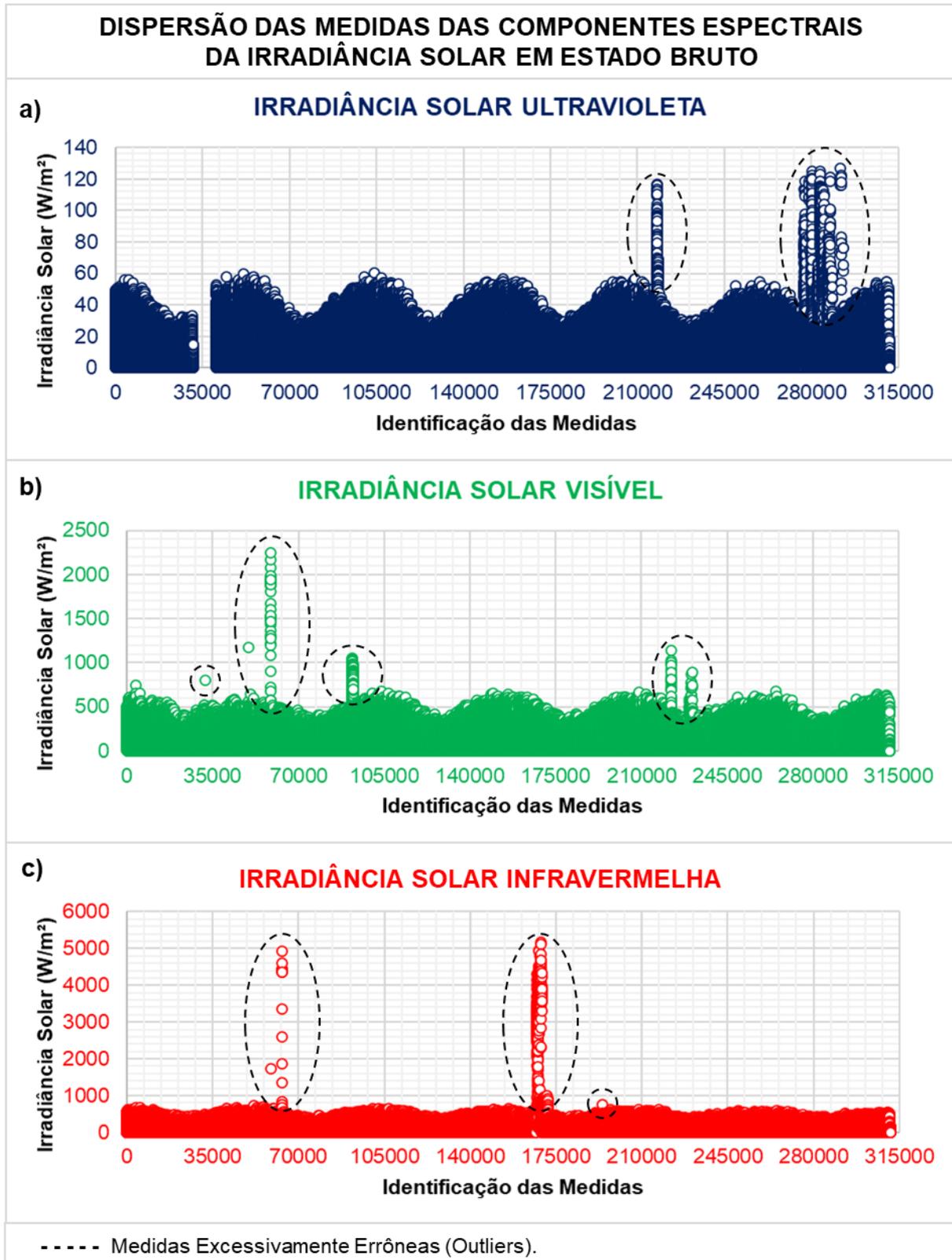
Tabela 9 - Quantidade total de medidas das irradiâncias solares ultravioleta, visível e infravermelha não-nulas sem validação em resolução temporal sub-horária.

ANO	IRRADIÂNCIA SOLAR		
	ULTRAVIOLETA	VÍVEL	INFRAVERMELHA
2001	42.301	51.704	50.474
2002	50.419	50.810	49.445
2003	51.861	51.873	50.622

2004	51.640	50.590	48.644
2005	51.809	50.851	50.369
2006	50.210	52.413	51.060
TOTAL	298.240	308.241	300.614

De acordo com Calça et al. (2019), é possível usar a inspeção visual da distribuição dos valores medidos em função do tempo como um procedimento para identificar medições errôneas e isoladas de irradiância solar, além dos princípios físicos e métodos estatísticos. No entanto, sua efetividade depende da experiência do responsável por sua aplicação, uma vez que é preciso ter conhecimento suficiente para identificar uma variação anormal nas medições. Neste estudo, a inspeção visual não foi utilizada para identificar medições errôneas ou isoladas, porém, através dela foi possível compreender, visualmente, que ambas as componentes espectrais da irradiância solar possuem valores excessivamente errôneos (outliers), pois se desviam demasiadamente dos demais na série temporal (Figura 10).

Figura 10 - Dispersão da série completa de irradiancias solares espectrais em estado bruto.



Segundo Pereira et al. (2017), valores mais elevados de irradiância solar são registrados durante o verão, devido ao solstício, que é o momento em que o Sol atinge

a maior declinação em latitude medida a partir do Equador, o que proporciona uma maior incidência solar. Assim como, no inverno, são registrados valores menores de irradiância solar, devido ao equinócio, que é o momento em que a duração do dia é igual à duração da noite em ambos os hemisférios, o que proporciona níveis menores de incidência solar. Portanto, se desviando desse padrão de comportamento, a irradiância solar ultravioleta (Figura 10a) apresentou duas concentrações de valores excessivos, próximas dos registros de número 210.000 e 280.000, caracterizando a menor ocorrência dentre as componentes espectrais. Já a irradiância solar visível (Figura 10b) apresentou quatro concentrações de valores discrepantes, próximas dos registros de número 35.000, 70.000, 105.000 e 210.000, a maior ocorrência em relação as demais componentes espectrais. Por fim, a irradiância solar infravermelha (Figura 10c) apresentou três concentrações de valores inconsistentes, próximas dos registros de número 70.000, 175.000 e 210.000.

4.2 Princípios Físicos de Detecção de Medidas Errôneas

De acordo com o primeiro princípio físico de detecção de medidas errôneas, onde são verificados se os valores estão dentro dos limites considerados fisicamente possíveis (Tabela 10), a irradiância solar ultravioleta não teve nenhum valor mensurado sinalizado como inválido. A irradiância solar visível teve a maior quantidade de medidas inválidas sinalizadas no ano de 2002, compreendendo 146 valores (0,287% do total para aquele ano), e teve o ano de 2001 com a menor quantidade, compreendendo 2 valores (0,004% do total para aquele ano), desconsiderando os anos sem sinalizações. A irradiância solar infravermelha teve o ano de 2004 com a maior quantidade de medidas inválidas sinalizadas, compreendendo 713 valores (1,466% do total para aquele ano), e a menor quantidade em 2002, compreendendo 11 valores (0,022% do total para aquele ano), desconsiderando os anos sem sinalizações.

Tabela 10 - Quantidade de medidas das irradiâncias solares sinalizadas como inválidas nas regras de limites fisicamente possíveis.

ANO	IRRADIÂNCIAS SOLAR					
	ULTRAVIOLETA		VÍSEL		INFRAVERMELHA	
	TOTAL	%	TOTAL	%	TOTAL	%
2001	0	0,000	2	0,004	0	0,000

2002	0	0,000	146	0,287	11	0,022
2003	0	0,000	0	0,000	0	0,000
2004	0	0,000	0	0,000	713	1,466
2005	0	0,000	38	0,074	0	0,000
2006	0	0,000	0	0,000	0	0,000

De acordo com o segundo princípio físico de detecção de medidas errôneas, onde são verificados se os valores estão dentro dos limites considerados extremamente raros (Tabela 11), a irradiância solar ultravioleta teve somente 140 valores sinalizados como inválidos no ano de 2006, representando 0,279% de todos os pontos medidos naquele ano. A irradiância solar visível teve a maior quantidade de medidas sinalizadas como inválidas no ano de 2002, assim como no princípio físico anterior, com 253 valores, representando 0,498% do total para aquele ano, e em 2004 teve a menor quantidade de medidas inválidas, desconsiderando o ano de 2006 que não teve sinalizações, compreendendo 1 valor sinalizado que representa 0,001% do total para aquele ano. A irradiância solar infravermelha teve no ano de 2004 a maior quantidade de medidas sinalizadas como inválidas, também como no princípio físico anterior, com 208 valores que representam 0,427% do total para aquele ano, de forma que os demais anos não receberam sinalizações de medições inválidas.

Tabela 11 - Quantidade de medidas das irradiâncias solares sinalizadas como inválidas nas regras de limites extremamente raros.

ANO	IRRADIÂNCIA SOLAR					
	ULTRAVIOLETA		VISÍVEL		INFRAVERMELHA	
	TOTAL	%	TOTAL	%	TOTAL	%
2001	0	0,000	16	0,030	0	0,000
2002	0	0,000	253	0,498	0	0,000
2003	0	0,000	2	0,003	0	0,000
2004	0	0,000	1	0,001	208	0,427
2005	0	0,000	41	0,080	0	0,000
2006	140	0,279	0	0,000	0	0,000

De acordo com o terceiro princípio físico de detecção de medidas errôneas, onde as medições são comparadas com os valores obtidos para o topo da atmosfera (Tabela 12), a irradiância solar ultravioleta teve no ano de 2006 a maior quantidade de medidas inválidas sinalizadas, compreendendo 1.447 valores (2,882% do total para aquele ano), e teve no ano de 2003 a menor quantidade de medidas inválidas, com somente 99 valores sinalizados (0,191% do total para aquele ano). A irradiância solar visível teve a maior quantidade de medidas inválidas sinalizadas no ano de 2002, com 368 valores que representam 0,724% do total para aquele ano, e em 2006 teve a menor quantidade de medidas sinalizadas como inválidas, compreendendo 221 valores, que representam 0,422% do total para aquele ano. A irradiância solar infravermelha teve o ano de 2004 com a maior quantidade de medidas inválidas sinalizadas, compreendendo 344 valores (0,707% do total para aquele ano), e teve em 2006 a menor ocorrência de medições inválidas, com 13 valores sinalizados (0,025% do total daquele ano).

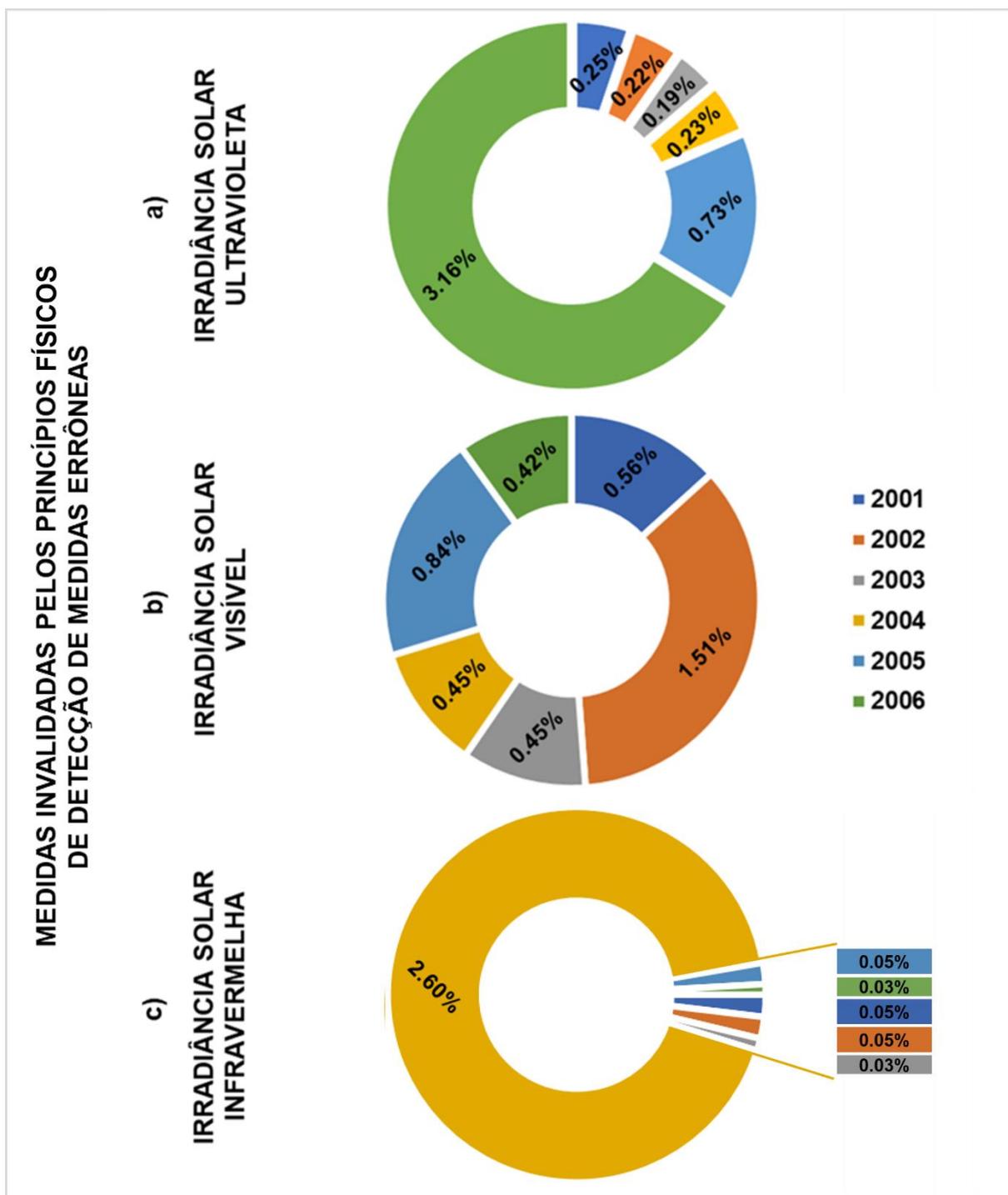
Tabela 12 - Quantidade de medidas das irradiâncias solares sinalizadas como inválidas nas regras de comparação com os valores no topo da atmosfera.

ANO	IRRADIÂNCIA SOLAR					
	ULTRAVIOLETA		VISÍVEL		INFRAVERMELHA	
	TOTAL	%	TOTAL	%	TOTAL	%
2001	105	0,248	273	0,528	27	0,053
2002	109	0,216	368	0,724	15	0,030
2003	99	0,191	234	0,451	15	0,030
2004	118	0,228	229	0,453	344	0,707
2005	379	0,731	349	0,686	26	0,052
2006	1.447	2,882	221	0,422	13	0,025

Considerando os três princípios físicos de detecção de medidas errôneas, em 2001, 2002, 2003 e 2005 a irradiância solar visível foi a que apresentou a maior quantidade de medidas sinalizadas como inválidas (Figura 11b), contabilizando 0,56%, 1,51%, 0,45% e 0,84% do total para cada ano, respectivamente. Tal qual, em 2004 a irradiância solar infravermelha foi a que apresentou a maior quantidade de medidas sinalizadas como inválidas (Figura 11c), contabilizando 2,60% do total para aquele ano. Bem como, no ano de 2006 a irradiância solar ultravioleta foi a que

apresentou a maior quantidade de medidas sinalizadas como inválidas (Figura 11a), contabilizando 3,16% do total para aquele ano. Em contrapartida, a irradiância solar infravermelha apresentou em 2001, 2002, 2003, 2005 e 2006 as menores quantidades de medições sinalizadas como inválidas, contabilizando 0,05%, 0,05%, 0,03%, 0,05% e 0,03% do total para cada ano, respectivamente. Assim como, a irradiância solar ultravioleta teve a menor quantidade de medições sinalizadas como inválidas em 2004, contabilizando 0,24% do total.

Figura 11 - Percentual de medidas invalidadas pelos princípios físicos do controle de qualidade.



Com relação aos princípios físicos de controle de qualidade, a maioria dos valores invalidados estavam acima do limite superior de variação. De forma que, os valores maiores que os limites fisicamente possíveis e extremamente raros indicam um erro excessivo de medição, causado por rompimento do cabeamento, má conexão dos fios no datalogger ou mal funcionamento dos sensores. Assim como, os valores

mensurados na superfície terrestre que estavam acima dos obtidos para o topo da atmosfera, em curtos períodos do dia, podem ter sido afetados pelo fenômeno da multirreflexão, relata Shi et al., (2008). Portanto, somente foram sinalizados como inválidos os valores que estavam 20% acima dos estimados para o topo da atmosfera (Equações 7, 8 e 9), de forma a indicar, corretamente, as medições errôneas.

4.3 Intervalo Estatístico de Detecção de Medidas Isoladas

Após a validação pelos princípios físicos de detecção de medidas errôneas, foram então encontrados valores estatísticos descritivos para cada componente espectral das irradiâncias solares, considerando o padrão de cobertura de céu aberto (Tabela 13a), parcialmente aberto (Tabela 13b), parcialmente nublado (Tabela 13c) e nublado (tabela 13d). Portanto, foram obtidas informações como a quantidade de pontos medidos, a média, o desvio padrão, o valor mínimo e o máximo para toda a série de irradiâncias solares. Para tanto, só se considerou medidas não sinalizadas como errôneas, que estavam de acordo com os valores limites impostos pelas três regras dos princípios físicos de controle de qualidade (Seção 3.3.1). Pôde-se perceber que a maior quantidade de valores foi obtida em condições de céu aberto ou nublado, e as menores quantias em condições de céu parcialmente nublado e parcialmente aberto, respectivamente, durante o período de 2001 a 2006.

Tabela 13 - Estatísticas descritivas das irradiâncias solares espectrais em função da cobertura de céu.

a)	COBERTURA DE CÉU ABERTO		
	IRRADIÂNCIA SOLAR (W/m ²)		
	ULTRAVIOLETA	VISÍVEL	INFRAVERMELHA
CONTAGEM	104608	108834	108187
MÉDIA	29,18963	334,04872	342,34825
DESVIO PADRÃO	9,77078	95,74068	90,93854
MÍNIMA	0,0064	0,0917	0,10249
MÁXIMA	87,1154	620,1835	781,88670
b)	COBERTURA DE CÉU PARCIALMENTE ABERTO		

ESTATÍSTICA	IRRADIÂNCIA SOLAR (W/m ²)		
	ULTRAVIOLETA	VISÍVEL	INFRAVERMELHA
CONTAGEM	35279	36725	36820
MÉDIA	17,81242	209,12615	217,4954
DESVIO PADRÃO	8,92601	83,92208	81,68152
MÍNIMA	0,109	0,7339	0,10249
MÁXIMA	85,2244	482,1101	719,1182
c)	COBERTURA DE CÉU PARCIALMENTE NUBLADO		
ESTATÍSTICA	IRRADIÂNCIA SOLAR (W/m ²)		
	ULTRAVIOLETA	VISÍVEL	INFRAVERMELHA
CONTAGEM	56436	58504	58463
MÉDIA	13,56325	146,05447	146,74505
DESVIO PADRÃO	8,96259	83,25886	81,54109
MÍNIMA	0,0128	0,0917	0,10249
MÁXIMA	84,0064	458,7156	760,13300
d)	COBERTURA DE CÉU NUBLADO		
ESTATÍSTICA	IRRADIÂNCIA SOLAR (W/m ²)		
	ULTRAVIOLETA	VISÍVEL	INFRAVERMELHA
CONTAGEM	97658	100322	94990
MÉDIA	6,40185	59,86361	55,52713
DESVIO PADRÃO	5,81836	52,03581	50,13332
MÍNIMA	0,0032	0,04590	0,052
MÁXIMA	79,6795	381,6514	501,5475

De acordo com o intervalo estatístico de detecção de medidas isoladas, onde as medições são comparadas com os valores da média, desvio padrão e nível de confiança de 99,70% (Tabela 14) em função da classificação de cobertura de céu, a irradiância solar ultravioleta teve no ano de 2006 a maior quantidade de medidas inválidas sinalizadas, compreendendo 264 valores (0,526% do total para aquele ano), e teve no ano de 2001 a menor quantidade de medidas inválidas, com somente 20

valores sinalizados (0,047% do total para aquele ano). A irradiância solar visível teve a maior quantidade de medidas inválidas sinalizadas no ano de 2003, com 71 valores que representam 0,137% do total para aquele ano, e em 2002 teve a menor quantidade de medidas sinalizadas como inválidas, compreendendo 28 valores, que representam 0,055% do total para aquele ano. A irradiância solar infravermelha teve no ano de 2004 com a maior quantidade de medidas inválidas sinalizadas, compreendendo 915 valores (1,881% do total para aquele ano), e teve em 2003 a menor ocorrência de valores medidos inválidos, com 46 valores sinalizados (0,091% do total daquele ano).

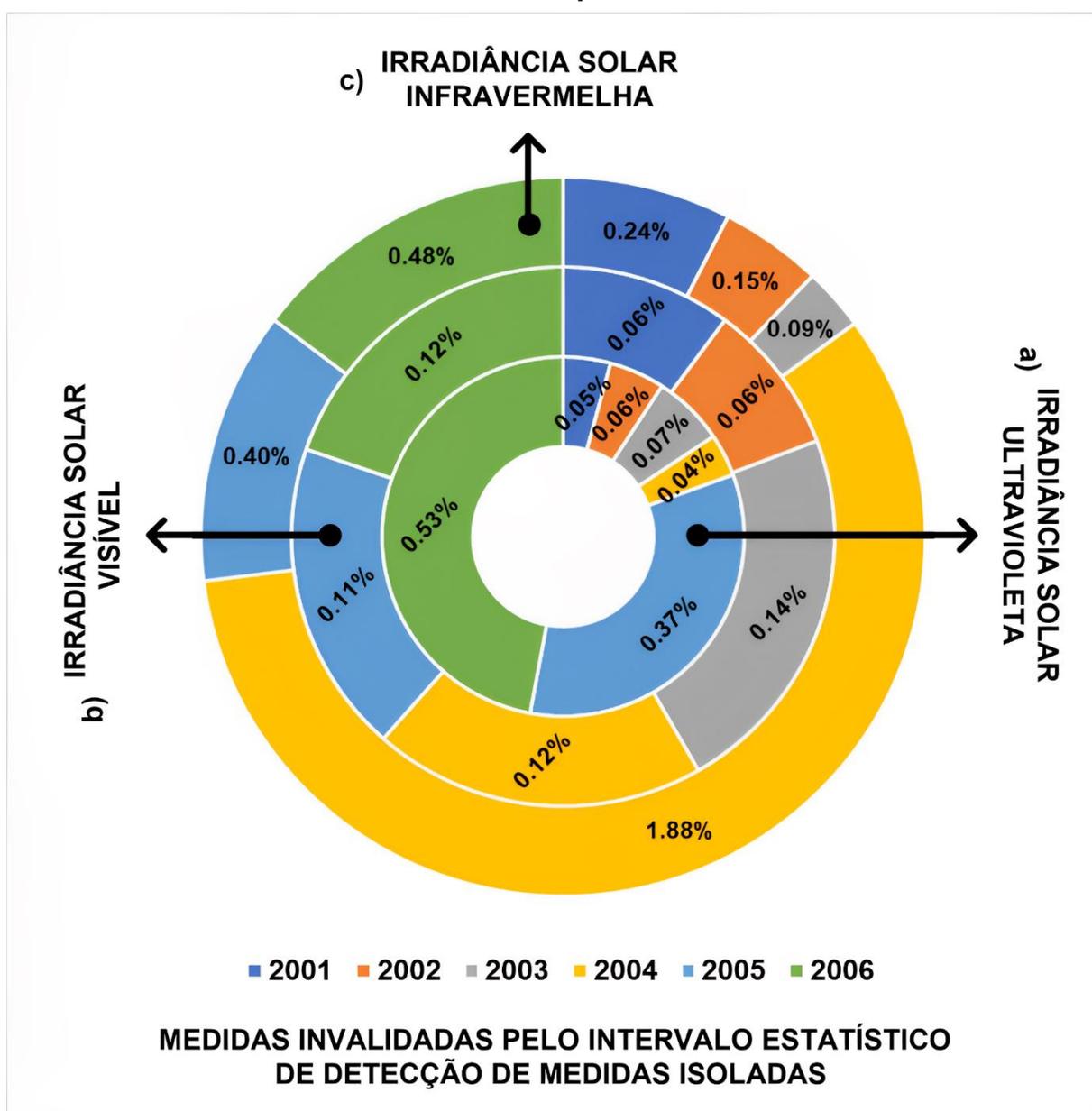
Tabela 14 - Quantidade de medidas das irradiâncias solares sinalizadas como inválidas pelos intervalos estatísticos de confiança.

ANO	IRRADIÂNCIA SOLAR					
	ULTRAVIOLETA		VISÍVEL		INFRAVERMELHA	
	TOTAL	%	TOTAL	%	TOTAL	%
2001	20	0,047	32	0,062	123	0,243
2002	28	0,055	28	0,055	73	0,148
2003	37	0,071	71	0,137	46	0,091
2004	22	0,042	61	0,120	915	1,881
2005	194	0,374	58	0,114	200	0,397
2006	264	0,526	63	0,120	243	0,476

Portanto, considerando o intervalo estatístico de detecção de medidas isoladas, em 2001, 2002, 2004 e 2005 a irradiância solar infravermelha foi a que apresentou a maior quantidade de medidas inválidas (Figura 12c), contabilizando 0,24%, 0,15%, 1,88% e 0,40% do total para cada ano, respectivamente. Assim como, em 2003 a irradiância solar visível foi a que apresentou a maior quantidade de medidas sinalizadas como inválidas (Figura 12b), contabilizando 0,14% do total para aquele ano. Tal que, no ano de 2006 a irradiância solar ultravioleta apresentou a maior quantidade de medidas sinalizadas como inválidas (Figura 12a), contabilizando 0,53% do total para aquele ano. Em contrapartida, em 2001, 2002, 2003 e 2004 a irradiância solar ultravioleta foi a que apresentou as menores proporções de medições invalidadas, totalizando 0,07%, 0,06%, 0,07% e 0,04% para cada ano, respectivamente. Em 2002, proporcionalmente similar a irradiância solar ultravioleta,

2005 e 2006 a irradiância solar visível apresentou as menores quantidades de medidas invalidadas, contabilizando 0,06%, 0,11% e 0,12% do total para cada ano, respectivamente.

Figura 12 - Percentual de medidas invalidadas pelo intervalo estatístico do controle de qualidade.



Em um contexto geral, em decorrência dos três princípios físicos e do intervalo estatístico de controle de qualidade (Tabela 15), a irradiância solar ultravioleta teve no ano de 2006 a maior quantidade de medidas inválidas sinalizadas, compreendendo 1.881 valores (3,746% do total para aquele ano), e teve no ano de 2001 a menor quantidade de medidas inválidas, com somente 125 valores sinalizados (0,295% do

total para aquele ano). A irradiância solar visível teve a maior quantidade de medidas inválidas sinalizadas no ano de 2002, com 795 valores que representam 1,565% do total para aquele ano, e em 2006 teve a menor quantidade de medidas sinalizadas como inválidas, compreendendo 284 valores, que representam 0,542% do total para aquele ano. A irradiância solar infravermelha teve o ano de 2004 com a maior quantidade de medidas inválidas sinalizadas, compreendendo 2.180 valores (4,481% do total para aquele ano), e teve em 2003 a menor ocorrência de valores medidos inválidos, com 61 valores sinalizados (0,120% do total daquele ano).

Tabela 15 - Quantidade de medidas das irradiâncias solares sinalizadas como inválidas por ano pelo procedimento de controle de qualidade.

ANO	IRRADIÂNCIA SOLAR					
	ULTRAVIOLETA		VISÍVEL		INFRAVERMELHA	
	TOTAL	%	TOTAL	%	TOTAL	%
2001	125	0,295	323	0,625	150	0,297
2002	137	0,271	795	1,565	99	0,200
2003	136	0,262	307	0,592	61	0,120
2004	140	0,271	291	0,575	2.180	4,481
2005	573	1,106	486	0,956	226	0,449
2006	1.881	3,746	284	0,542	256	0,501

Considerando o primeiro princípio físico de detecção de medidas errôneas (limites fisicamente possíveis), a componente espectral infravermelha teve a maior e a ultravioleta a menor quantidade de medidas sinalizadas como inválidas, representando 0,241% e 0,000% do total, respectivamente. Assim como, considerando o segundo princípio físico de detecção de medidas errôneas (limites extremamente raros), a componente espectral visível teve a maior e a ultravioleta a menor quantidade de medidas sinalizadas como inválidas, representando 0,101% e 0,046% do total, respectivamente. Bem como, considerando o terceiro princípio físico de detecção de medidas errôneas (valores para o topo da atmosfera), a componente espectral ultravioleta teve a maior e a infravermelha a menor quantidade de medidas sinalizadas como inválidas, compreendendo 0,767% e 0,146% do total, respectivamente. Por fim, levando em consideração o intervalo estatístico de detecção de medidas isoladas, a componente espectral infravermelha foi a que teve a maior e

a visível a menor quantidade de medidas sinalizadas como inválidas, compreendendo 0,101% e 0,532% do total, respectivamente (Tabela 16).

Tabela 16 - Quantidade final de valores das irradiâncias solares sinalizadas como inválidas após o procedimento de controle de qualidade.

CONTROLE DE QUALIDADE	IRRADIÂNCIA SOLAR					
	ULTRAVIOLETA		VISÍVEL		INFRAVERMELHA	
	TOTAL	%	TOTAL	%	TOTAL	%
TOTAL DE MEDIDAS	298.240	100,0	308.241	100,00	300.614	100,00
1º PRINCÍPIO FÍSICO	0	0,000	186	0,060	724	0,241
2º PRINCÍPIO FÍSICO	140	0,046	313	0,101	208	0,069
3º PRINCÍPIO FÍSICO	2.287	0,767	1.674	0,543	440	0,146
INTERVALOS ESTATÍSTICOS	565	0,189	313	0,101	1.600	0,532
TOTAL INVÁLIDO	2.992	1,003	2.486	0,806	2.972	0,989

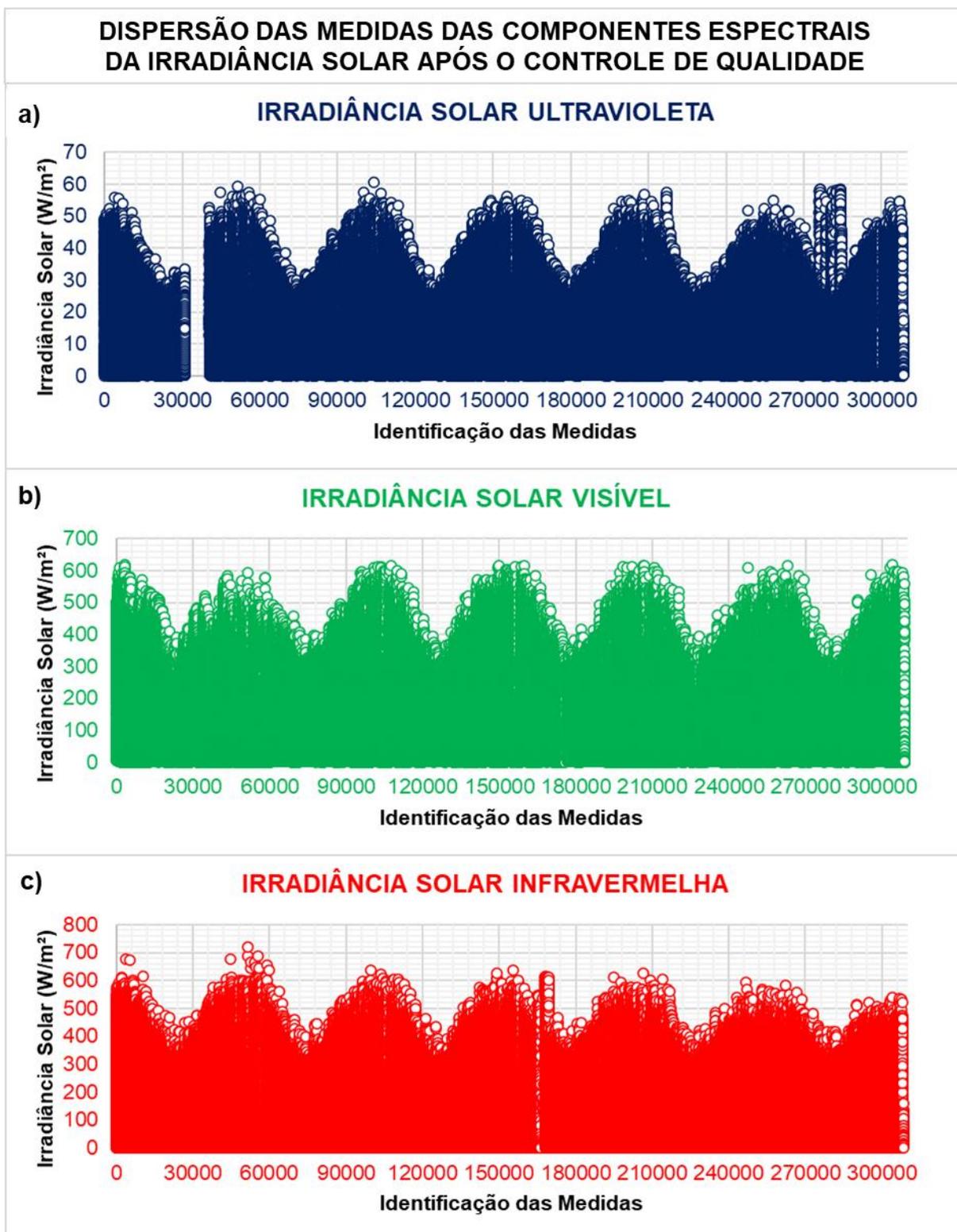
Legenda: **Total de Medidas** - Quantidade total de valores mensurados de 2001 a 2006; **1º Princípio Físico** - Regras de análise dos limites fisicamente possíveis; **2º Princípio Físico** - Regras de análise dos limites extremamente raros; **3º Princípio Físico** - Regras de comparação com os valores obtidos para o topo da atmosfera; **Intervalos Estatísticos** - Regras de análise das medições com relação aos intervalos estatísticos de confiança; e **Total Inválido** - Quantidade total de valores não-nulos sinalizados como inválidos de 2001 a 2006 no procedimento de controle de qualidade.

A irradiância solar ultravioleta foi a componente espectral que apresentou a maior quantidade de medidas inválidas no período de 2001 a 2006, com 2.992 valores sinalizados como errôneos ou isolados, representando 1,003% do total (Tabela 16). Em seguida, se tem a irradiância solar infravermelha com 2.972 valores sinalizados como inválidos por estarem em desacordo com os princípios definidos (princípios físicos e intervalo estatístico), que representam 0,989% do total. Por fim, a irradiância solar visível foi a que teve a menor quantidade de medidas sinalizadas como errôneas ou isoladas, com 2.486 valores invalidados, que representam 0,806% do total. Todas as medidas sinalizadas como inválidas podem ser consideradas outliers, pois estão

incorretas com relação aos valores teóricos pré-definidos (princípios físicos) ou se afastam significativamente dos limites definidos pelo intervalo estatístico de 99,70% de confiança.

Em relação a dispersão da série de medidas das irradiâncias solares após a aplicação do procedimento de controle de qualidade, a componente espectral visível foi a que apresentou o melhor padrão de distribuição dos pontos de medições desconsiderando os outliers (Figura 13b). Tal que, para a componente espectral infravermelha foi possível identificar que os princípios físicos e o intervalo estatístico reduziu a maior parte dos outliers, porém alguns ainda podem ser identificados visualmente (Figura 13c), próximos dos registros de número 60.000 e 150.000. Por fim, constatou-se também que os outliers da componente espectral ultravioleta foram identificados em sua maior parte, no entanto foi observado, a partir da inspeção visual, uma porção de pontos de medições distribuídos de forma errônea, próxima dos registros de número 210.000 e, principalmente do 270.000, com uma ocorrência mais dúbia em relação a qualidade dos valores mensurados (Figura 13a).

Figura 13 - Dispersão da série de irradiâncias solares espectrais após o controle de qualidade.



Uma forma de se obter um melhor resultado ao identificar medições errôneas ou isoladas, em relação aos pontos medidos que não foram detectados pelo procedimento de controle de qualidade proposto (Figura 13), é restringir ainda mais

as regras de comparação com os valores obtidos para o topo da atmosfera a partir do terceiro princípio físico (Equações 7, 8 e 9). Outra forma, que deve ser averiguada com cautela, seria restringir, também, o nível de confiança do intervalo estatístico, de acordo com a distribuição normal de probabilidade (Tabela 5). No entanto, como mencionado anteriormente, realizar uma inspeção visual da distribuição da série temporal de medidas também permitiria identificar valores errôneos ou isolados, porém o processo de sinalização dessas medições seria mais laborioso, uma vez que a maioria das ferramentas computacionais de análise de dados permite somente eliminar esses pontos medidos da série temporal, e não os sinalizar como inválidos.

Torna-se importante destacar também que não existem informações detalhadas sobre os problemas que acometeram o sistema de medição durante os anos de 2001 a 2006. Apesar de existir a instrumentação e a coleta eletrônica dos valores das irradiâncias solares espectrais, não existia um sistema computacional de gerenciamento das informações geradas no período de execução do experimento. Problemas técnicos relacionados a alimentação elétrica dos aparelhos, mal funcionamento dos sensores e do datalogger foram relatados por Escobedo et al. (2011), com maior ênfase para os anos de 2004 e 2005, onde os autores desconsideraram 80 e 35 dias de medidas em seu estudo, respectivamente, e com menores ocorrências para os anos de 2001, 2002 e 2003, com 1, 15 e 6 dias também desconsiderados para cada ano, respectivamente, retirando do conjunto de medidas valores obtidos das três componentes espectrais.

4.4 Preenchimento de Lacunas com Machine Learning

Com relação a todos os algoritmos de Machine Learning experimentados (Seção 3.4), o que gerou o melhor resultado em relação as métricas de avaliação para as componentes espectrais da irradiância solar foi o Light GBM (2022). O modelo de estimativa da componente espectral infravermelha foi o que obteve as melhores métricas em relação aos das componentes espectrais ultravioleta e visível, com coeficiente de determinação (R^2) de 0,99137, erro médio absoluto relativo (rMAE) de 3,923% e raiz quadrada do erro médio relativo (rRMSE) de 6,889%. Neste sentido, o modelo de estimativa das componentes espectrais visível e ultravioleta obtiveram o segundo e o terceiro melhor resultado, com R^2 de 0,99562 e 0,99137, rMAE de 2,886% e 3,923% e rRMSE de 4,875% e 6,889%, respectivamente (Tabela 17).

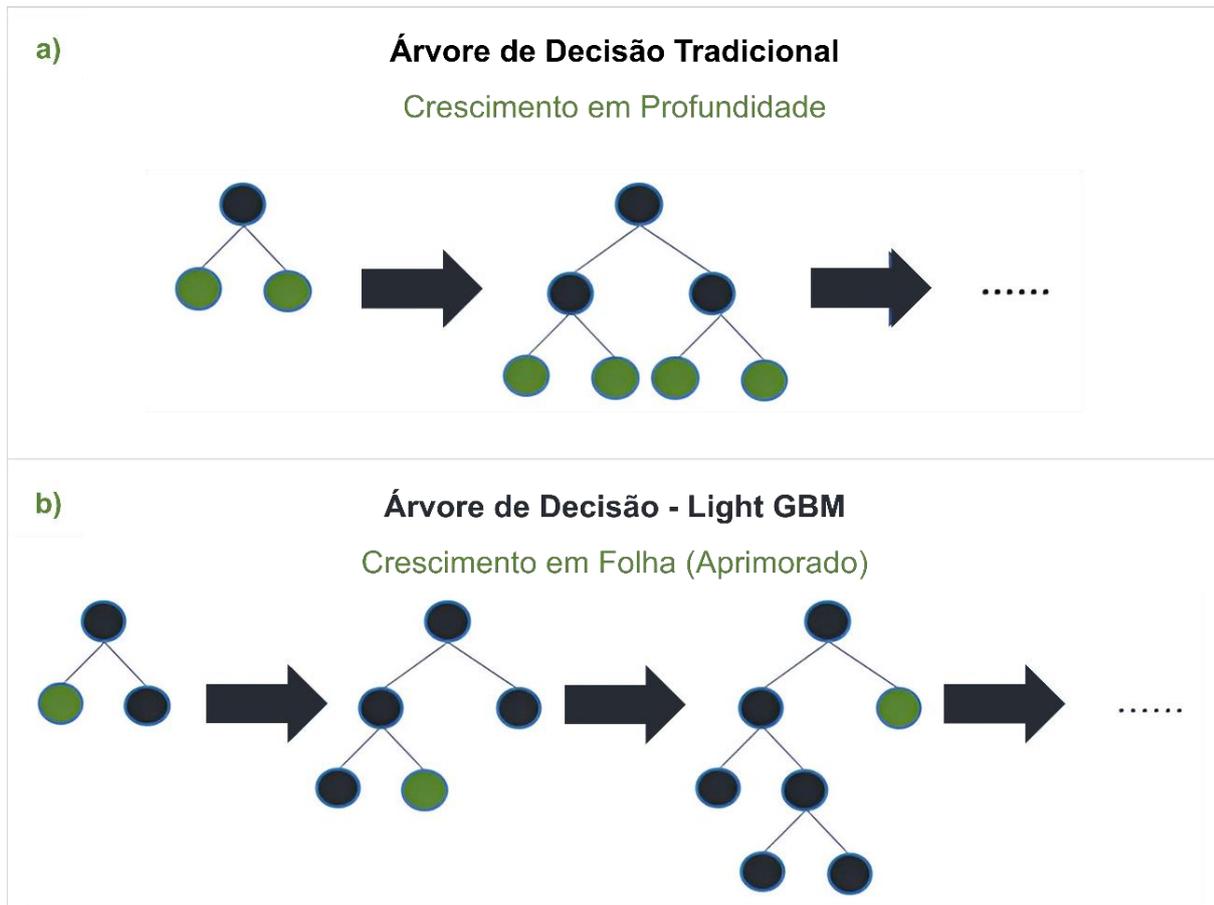
Tabela 17 - Métricas de avaliação dos modelos Machine Learning das irradiâncias solares ultravioleta, visível e infravermelha.

MODELO DE IRRADIÂNCIA SOLAR LIGHT GBM	MAE (W/m²)	rMAE (%)	RMSE (W/m²)	rRMSE (%)	R²
Ultravioleta	0,67025	3.923	1,17685	6.889	0,99137
Visível	5,50889	2.886	9,30457	4.875	0,99562
Infravermelha	5,54053	2.849	9,37623	4.821	0,99565

Pode-se apontar dois motivos para o modelo de Machine Learning de estimativa da irradiância solar ultravioleta ter tido o pior resultado, quando comparado aos das outras duas componentes espectrais (visível e infravermelha). O primeiro motivo, é a lacuna observada próxima ao registro de número 30.000 na série temporal (Figura 13b). O segundo motivo é a quantidade de medições visualmente errôneas e isoladas, que não foram sinalizadas por nenhum dos métodos do procedimento de controle de qualidade, próxima do registro de número 270.000 (Figura 13b). Neste caso, a ausência de valores ou a entrada de parâmetros inconsistentes no processo de treinamento de um modelo de Machine Learning podem gerar resultados inferiores quando comparados aos gerados por uma série temporal inteira.

O Light GBM é um algoritmo de Machine Learning baseado em aprimoramento de Árvore de Decisão (Gradient Boosting), com histogramas que agrupam valores de recursos contínuos em caixas discretas e aprendizado paralelo que interpreta várias camadas concomitantemente (SCIKIT-LEARN, 2022b). O que possibilita que o treinamento de modelos seja acelerado e reduza o uso de memória computacional. De acordo com KE et al. (2017), diferente dos demais algoritmos de Árvore de Decisão, que tem o crescimento em profundidade, ou seja, as camadas se ampliam formando mais ramificações para baixo (Figura 14a), ele tem seu crescimento em folha, onde a camada lateral que gera o melhor resultado é a que se expande, pois tende a gerar uma saída mais aprimorada (Figura 14b).

Figura 14 - Estrutura do algoritmo de Gradient Boosting (Machine Learning).



Fonte: Adaptado de Light GBM (2022).

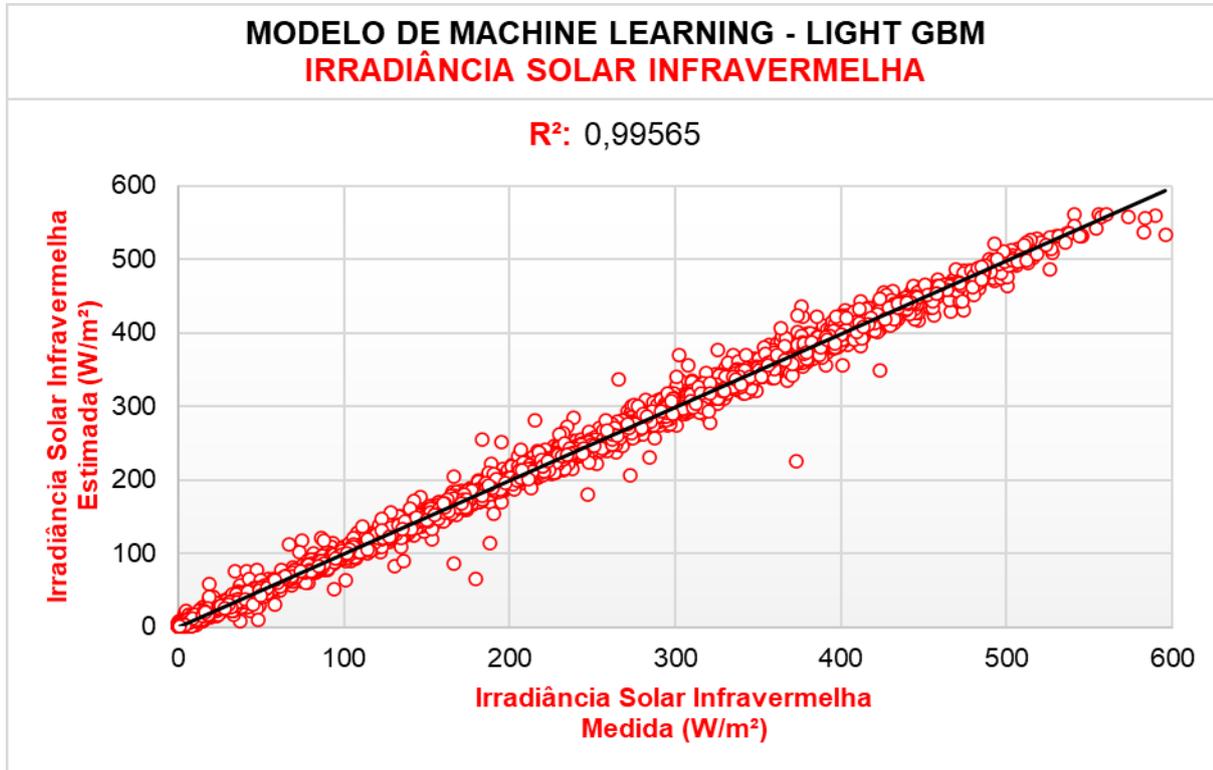
Para que se possa aferir se as métricas e gráficos foram gerados corretamente, assim como explorar melhor os resultados obtidos, o Azure Machine Learning fornece uma amostra de saída contendo cinco mil linhas, separadas aleatoriamente, que compreendem os valores verdadeiros (medidos) e os modelados (estimados) para cada componente espectral da irradiância solar. Portanto, para fins de averiguação dos resultados apresentados, as três amostras geradas das irradiâncias solares ultravioleta (bit.ly/ultraviolet-sample), visível (bit.ly/visible-sample) e infravermelha (bit.ly/infrared-sample), foram disponibilizadas via Git Hub em um repositório público. É importante destacar também que, optou-se por exportar essas amostras a fim de gerar gráficos mais aprimorados, fora da Microsoft Azure.

Os resultados dos modelos de Machine Learning foram comparados com os resultados obtidos a partir da modelagem de Escobedo et al. (2011), baseada na mesma série temporal de medições. Após quatro anos decorridos do período de coleta

das medidas, Escobedo et al. (2011), responsáveis pela condução do experimento que durou seis anos no Laboratório de Agrometeorologia e Radiometria Solar da Faculdade de Ciências Agronômicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu (SP) - Brasil, exploraram a evolução sazonal e ajustaram modelos de Regressão Linear Simples para estimativa das componentes espectrais ultravioleta, visível e infravermelha em função das medidas da irradiância solar global, em resolução temporal horária e diária.

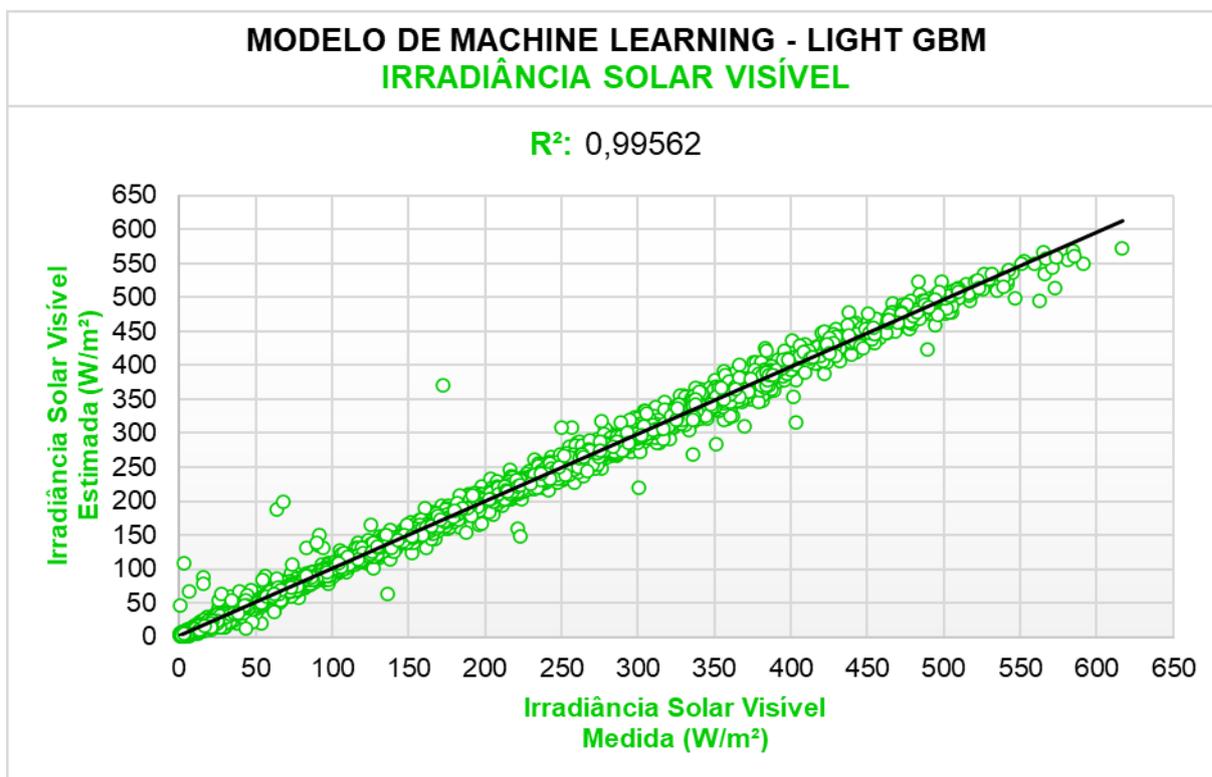
O modelo de Machine Learning da irradiância solar infravermelha produzido (Figura 15) gerou uma tendência em superestimar os valores sub-horários em 4,821% (rRMSE). Escobedo et al. (2011), obtiveram também uma tendência em superestimar, porém menor, em seu modelo de Regressão Linear Simples, se comparado ao deste estudo, com rRMSE de 3,73% para os valores diários e maior com rRMSE de 5,27% para os valores horários da componente espectral infravermelha. Com relação ao coeficiente de determinação, Escobedo et al. (2011) obtiveram R^2 de 0,9968 em resolução temporal horária e R^2 de 0,9870 em resolução temporal diária, o que indica que o coeficiente de determinação horário foi maior e o diário menor, quando comparados aos obtidos como resultado deste estudo (R^2 de 0,99565) em resolução temporal sub-horária.

Figura 15 - Resultado do modelo de Machine Learning da irradiância solar infravermelha.



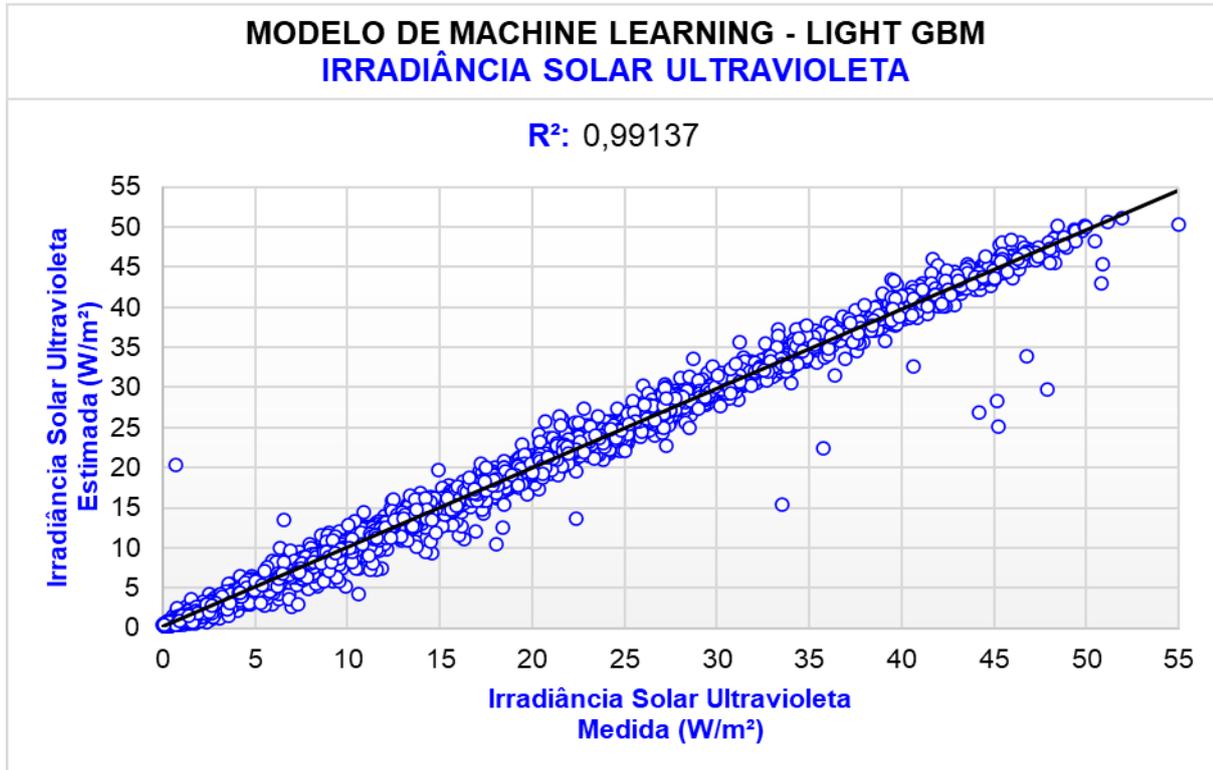
O modelo de Machine Learning da irradiância solar visível produzido (Figura 16) gerou uma tendência em superestimar os valores sub-horários em 4,875% (rRMSE). Escobedo et al. (2011), obtiveram também uma tendência em superestimar, porém menor, em seus modelos de Regressão Linear Simples, se comparados as métricas obtidas neste estudo, com rRMSE de 2,96% para os valores diários e rRMSE de 4,46% para os valores horários da componente espectral visível. Com relação ao coeficiente de determinação, Escobedo et al. (2011) obtiveram R^2 de 0,9978 em resolução temporal horária e R^2 de 0,9904 em resolução temporal diária, o que indica que o coeficiente de determinação horário foi maior e o diário menor quando comparados ao obtido como resultado deste estudo (R^2 de 0,99562) em resolução temporal sub-horária.

Figura 16 - Resultado do modelo de Machine Learning da irradiância solar visível.



Por fim, o modelo de Machine Learning da irradiância solar ultravioleta produzido (Figura 17) gerou uma tendência em superestimar os valores sub-horários em 6,889% (rRMSE). Escobedo et al. (2011), obtiveram uma tendência maior em superestimar em seus modelos de Regressão Linear Simples, se comparados às métricas obtidas neste estudo, com rRMSE de 9,94% para os valores diários e rRMSE de 8,39% para os valores horários da componente espectral ultravioleta. Com relação ao coeficiente de determinação, Escobedo et al. (2011) obtiveram R² de 0,9794 em resolução temporal horária e R² de 0,9204 em resolução temporal diária, ambos menores quando comparados ao coeficiente de determinação obtido como resultado deste estudo, com R² de 0,99137 em resolução temporal sub-horária.

Figura 17 - Resultado do modelo de Machine Learning da irradiância solar ultravioleta.



O resultado menor de rRMSE gerado pelos modelos sub-horários de estimativa das irradiâncias solares visível e infravermelha deste estudo, podem ser explicados pela grande variabilidade que se tem em valores medidos a cada 5 minutos, onde se está mais suscetível às adversidades relacionadas a instrumentação e as condições de operação do sistema de medição. Visto que, quando se gera valores para a resolução temporal horária e diária esses problemas tendem a ser reduzidos pelos métodos de integração e média numérica, o que reduz o impacto no resultado obtido. No entanto, apesar da diferença dos resultados encontrados quando comparados aos de Escobedo et al. (2011), pode-se perceber a mesma tendência em superestimar os valores em resolução temporal sub-horária, horária e diária.

Após a interpretação dos resultados apresentados, decidiu-se implantar os modelos em formato de serviço consumido pela internet (Web Service). Neste sentido, para cada modelo de estimativa, da irradiância solar ultravioleta, visível e infravermelha, foi disponibilizado um código para consumo da API Rest em linguagem de programação R (bit.ly/api-code-r) ou Python (bit.ly/api-code-python) via repositório público do Git Hub. Além dos códigos de consumo das APIs em Python e R, o usuário também precisará informar a URL e a Chave de Autenticação da API Rest do Web

Service da componente ultravioleta (bit.ly/ultraviolet-web-service-connect), visível (bit.ly/visible-web-service-connect) e infravermelha (bit.ly/infrared-web-service-connect), pois sem essas informações os códigos implementados não funcionam.

5 CONCLUSÕES

O procedimento de detecção de valores errôneos e isolados implementado na série temporal de medidas das componentes espectrais da irradiância solar (2001 a 2006) do Laboratório de Agrometeorologia e Radiometria Solar da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu (SP) - Brasil, permitiu obter as seguintes conclusões:

- a)** Após a aplicação do primeiro princípio físico de controle de qualidade, responsável por detectar medidas fisicamente possíveis, foi possível identificar que a irradiância solar infravermelha foi a que teve a maior e a ultravioleta a menor quantidade de valores sinalizados como inválidos.
- b)** Após a aplicação do segundo princípio físico de controle de qualidade, responsável por detectar medidas extremamente raras, foi possível identificar que a irradiância solar visível foi a que teve a maior e a ultravioleta a menor quantidade de valores sinalizados como inválidos.
- c)** Após a aplicação do terceiro princípio físico de controle de qualidade, responsável por comparar valores medidos com os do topo da atmosfera, foi possível identificar que a irradiância solar ultravioleta foi a que teve a maior e a infravermelha a menor quantidade de valores invalidados.
- d)** Após a aplicação do intervalo estatístico em função do padrão de cobertura de céu, responsável por detectar medidas isoladas, foi possível identificar que a irradiância solar infravermelha foi a que teve a maior e a visível a menor quantidade de valores sinalizados como inválidos.

Foi possível identificar também que, em um contexto geral, o terceiro princípio físico foi o mais e o primeiro o menos restritivo do procedimento de controle de qualidade. Neste sentido, a irradiância solar visível foi a que teve a menor e a ultravioleta a maior quantidade de medidas sinalizadas como inválidas pelos princípios físicos e intervalo estatístico de detecção de medidas errôneas e isoladas.

A criação do modelo de Machine Learning, na Microsoft Azure, gerou resultados excelentes, principalmente ao se considerar a variabilidade que se tem em valores sub-horários das irradiâncias solares espectrais. Tal que, as métricas obtidas, inclusive, se assemelham as de Escobedo et al. (2011) em modelagens horárias e diárias produzidas usando a mesma série de medições. Para estudos futuros sugere-se a utilização de um método de Machine Learning para a classificação de medições boas e ruins das componentes espectrais da irradiância solar, uma vez que facilitaria

o processo de controle de qualidade para profissionais que não possuem conhecimento técnico na área, assim como não possuem ferramentas computacionais disponíveis para trabalhar com grandes volumes de dados.

REFERÊNCIAS

- ALBERTAZZI, A.; SOUSA, A. R. **Fundamentos da Metrologia Científica e Industrial**. 2. ed. Barueri: Manole, 2018.
- ARENALES, S.; DAREZZO, A. **Cálculo Numérico: Aprendizagem com Apoio de Software**. São Paulo: Thomson Learning, 2008.
- AGGARWAL, S. K.; SAINI, L. M. Solar energy prediction using linear and non-linear regularization models: a study on AMS (American Meteorological Society) 2013-14 solar energy prediction contest. **Energy**, v. 78, p. 247-256, 2014.
- BONAFARI, F. C. **Estatística Inferencial**. São Paulo: Pearson Education do Brasil, p. 103-134, 2012.
- CALÇA, M. V. C. et al. Outliers Detection in a Quality Control Procedure for Measurements of Solar Radiation. **IEEE Latin America Transactions**, v. 17, n. 11, p. 1815-1822, 2019.
- CAMPBELL SCIENTIFIC. **SM 192/716 Storage Module**, Instruction Manual. 1993. Disponível em: <https://s.campbellsci.com/documents/us/manuals/sm192716.pdf>. Acesso em: 12 jan. 2022.
- CAMPBELL SCIENTIFIC. **PC208W - Datalogger Support Software for Windows**, Brochure. 2001. Disponível em: https://s.campbellsci.com/documents/us/product-brochures/b_pc208w.pdf. Acesso em: 12 de Janeiro de 2022.
- CAMPBELL SCIENTIFIC. **CR23X Micrologger**, Operator's Manual. 2006. Disponível em: <https://s.campbellsci.com/documents/us/manuals/cr23x.pdf>. Acesso em: 12 de Janeiro de 2022.
- CARRILHO, A. C.; GALO, M. **Remoção de Pontos Espúrios em Dados LiDAR Aerotransportado a Partir da Análise Estatística das Altitudes**. In: IVSBG Annual Meeting, 2017. Disponível em: http://docs.fct.unesp.br/departamentos/cartografia/eventos/2017_IV_SBG/_artigos/2017_SBG_CTIG_paper_92.pdf. Acesso em: 11 fev. 2022.
- CHARUCHITIPAN, D. et al. A semi-empirical model for estimating diffuse solar near infrared radiation in Thailand using ground- and satellite-based data for mapping applications. **Renewable Energy**, v. 117, p. 175-183, 2018.
- CORRÊA, F. H. P. **Estimativa das radiações ultravioleta (UV), fotossinteticamente ativa (PAR) e infravermelha (IV) em função da razão de insolação**. Dissertação (Mestrado em Agronomia/Energia na Agricultura) - Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista (UNESP), Câmpus de Botucatu, 2011.
- DRIEMEL, A. et al. Baseline Surface Radiation Network (BSRN): structure and data description (1992-2017). **Earth System Science Data**, v. 10, p. 1491-1501, 2018.

EPPLEY. **Eppley PSP - Precision Spectral Pyranometer**, Application Note. 2001. Disponível em: <https://s.campbellsci.com/documents/au/technical-papers/epp-ppsp.pdf>. Acesso em: 12 de Janeiro de 2022.

ESCOBEDO, J. F. et al. Radiações solares UV, PAR e IV: I - Estimativa em função da global. **Avances en Energías Renovables y Medio Ambiente**, v. 10, p. 79-86, 2006.

ESCOBEDO, J. F. et al. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. **Applied Energy**, v. 86, p. 299-309, 2009.

ESCOBEDO, J. G. et al. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. **Renewable Energy**, v. 36, p. 169-178, 2011.

ESTAÇÃO METEOROLÓGICA LAGEADO. **Laboratório de Agrometeorologia e Radiometria Solar**, Faculdade de Ciências Agrônomicas, Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, Botucatu (SP). 2022. Disponível em: <https://estacaolageado.fca.unesp.br/>. Acesso em: 14 de Janeiro de 2022.

GALLO, R. et al. Solar radiation forecasting with deep learning techniques integrating geostationary satellite images. **Engineering Applications of Artificial Intelligence**, v. 116, p. 105493, 2022.

GUEYMARD, C. A.; MYERS, D. R. Solar Radiation Measurement: Progress in Radiometry for Improved Modeling. **Modeling Solar Radiation at the Earth Surface**, p. 1-27, 2008.

GUEYMARD, C. A.; RUIZ-ARIAS, J. A. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global Irradiance. **Solar Energy**, v. 128, p. 1-30, 2016.

IQBAL, M. **An Introduction to Solar Radiation**. Ontario: Academic Press Canada, 1983.

INPE. **Sistema de Organização Nacional de Dados Ambientais**. 2018. Disponível em: <http://sonda.ccst.inpe.br/infos/validacao.html>. Acesso em: 17 jan. 2019.

JANJAI, S.; WATTAN, R.; SRIPRADIT, A. Modeling the ratio of photosynthetically active radiation to broadband global solar radiation using ground and satellite-based data in the tropics. **Advances in Space Research**, v. 56, n. 11, 2015.

JOURNÉE, M.; BERTRAND, C. Quality control of solar radiation data within the RMIB solar measurements network. **Solar Energy**, v. 85, p. 72-86, 2011.

KE, G. et al. **LightGBM: A Highly Efficient Gradient Boosting Decision Tree**. In: 31st Conference on Neural Information Processing Systems - NIPS, Long Beach - CA, USA. 2017. Disponível em:

<https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>. Acesso em: 16 nov. 2022.

KIPP & ZONEN. **CUV3 - Broadband UV Radiometer**, Instruction Manual. 2004. Disponível em: <https://www.kippzonen.com/Download/94/CUV-3-Manual?ShowInfo=true>. Acesso em: 12 de Janeiro de 2022.

KIPP & ZONEN. **PQS 1 PAR Quantum Sensor: For Measuring Photosynthetically Active Radiation**. Holanda: Kipp & Zonen, 2022. Disponível em: <https://www.kippzonen.com/Download/429/PQS-1-PAR-Quantum-Sensor-Brochure>. Acesso em: 24 jan. 2022.

KIPP & ZONEN. **Broadband UV Radiometers: For the Accurate Measurement of Solar Ultraviolet Radiation**. Holanda: Kipp & Zonen, 2022. Disponível em: <https://www.kippzonen.com/Product/163/CUV5-Total-UV-Radiometer#.YgL9QJbMLIU>. Acesso em: 08 fev. 2022.

LAMNATOU C.; CHEMISANA D. Solar radiation manipulations and their role in greenhouse claddings: fresnel lenses, NIR- and UV-blocking materials. **Renewable and Sustainable Energy Reviews**, v. 18, p. 271-287, 2013.

LECKNER, B. The Spectral Distribution of Solar Radiation at The Earth's Surface-Elements of a Model. **Solar Energy**, v. 20, p. 143-150, 1978.

LIGHT GMB. **Features**. Documentation. 2022. Disponível em: <https://lightgbm.readthedocs.io/en/v3.3.2/Features.html#references>. Acesso em 16 nov. 2022.

LI-COR BIOSCIENCES. **LI-190SA Quantum Sensor**. Lincoln: LI-COR Biosciences, 2022. Disponível em: <https://www.licor.com/env/pdf/light/190.pdf>. Acesso em: 25 jan. 2023.

LONG, C. N.; DUTTON, E. G. **BSRN Global Network Recommended Quality Control Tests**. BSRN Technical Report, v. 2, 2002. Disponível em: https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf. Acesso em: 26 out. 2018.

MACHADO, A. B.; RAMPINELLI, G. A. **Avaliação de métodos probabilísticos para preenchimento de lacunas de dados solarimétricos e previsão da irradiância solar de curto prazo**. In: VII Congresso Brasileiro de Energia Solar, Gramado - RS, Brasil. 2018.

MICROSOFT AZURE. **Azure Machine Learning**. Product. 2022a. Disponível em: <https://azure.microsoft.com/en-us/products/machine-learning/#product-overview>. Acesso em: 15 nov. 2022.

MICROSOFT AZURE. **Whats is machine learning platform**. Cloud Computing Dictionary. 2022b. Disponível em: <https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/what-is-machine-learning-platform/>. Acesso em: 26 set. 2022.

MICROSOFT LEARN. **Azure Machine Learning Documentation**. Documentation. 2022a. Disponível em: <https://learn.microsoft.com/en-us/azure/machine-learning/>. Acesso em: 26 set. 2022.

MICROSOFT LEARN. **Configure Training, Validation, Cross-Validation and Test Data in Automated Machine Learning**. Documentation. 2022b. Disponível em: <https://learn.microsoft.com/en-us/azure/machine-learning/how-to-configure-cross-validation-data-splits>. Acesso em: 26 set. 2022.

MORADI, I. Quality control of global solar radiation using sunshine durations hours. **Energy**, v. 34, p. 1-6, 2009.

MUNEER, T.; YOUNES, S.; MUNAWWAR. Discourses on solar radiation modeling. **Renewable and Sustainable Energy**, v. 11, p. 551-602, 2007.

NREL. **User's Manual for SERI QC Software: Assessing the Quality of Solar Radiation Data**. Golden: National Renewable Energy Laboratory, 1993.

NWOKOLO, S. C.; OGBULEZIE, J. C.; OBWULU, A. U. Impacts of climate change and meteo-solar parameters on photosynthetically active radiation prediction using hybrid machine learning with Physics-based models. **Advances in Space Research**, v. 70, n. 11, 2022.

OHMURA, A. et al. Baseline Surface Radiation Network (BSRN/WCRPP): New Precision Radiometry for Climate Research. **Bulletin of the American Meteorological Society**, v. 79, p. 2115-2136, 1998.

PAGAEAE. **BSRN Toolbox**. 2019. Disponível em: https://wiki.pangaea.de/wiki/BSRN_Toolbox#Quality_Check. Acesso em: 10 set. 2019.

PAGE J. **Proposed quality control procedures for the Meteorological Office data tapes relating to global solar radiation, diffuse solar radiation, sunshine, and cloud in the UK**. Report to CIBSE Guide Solar Data Task Group, Chartered Institute of Building Services Engineers, London, UK, 1997.

PASHIARDIS, S.; KALOGIROU, S. A. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. **Renewable Energy**, v. 96, p. 1015-1033, 2016.

PEREIRA, E. B. et al. **Atlas Brasileiro de Energia Solar**. 2 ed., São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2017.

QIN, W. et al. A novel efficient broadband model to derive daily surface solar Ultraviolet radiation (0.280–0.400 μm). **Science of the Total Environment**, v. 735, p. 139513, 2020.

REIKARD, G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. **Solar Energy**, v. 83, p. 342-349, 2009.

ROSSI, T. J. et al. Global, diffuse and direct solar radiation of the infrared spectrum in Botucatu / SP / Brazil. **Renewable and Sustainable Energy Reviews**, v. 82, p. 448-459, 2018.

SCIKIT-LEARN. **Model Evaluation**. 2022a. Documentation. Disponível em: https://scikit-learn.org/0.22/modules/model_evaluation.html#mean-absolute-error. Acesso em: 13 nov. 2022.

SCIKIT-LEARN. **Supervised Learning**. 2022b. Documentation. Disponível em: https://scikit-learn.org/stable/supervised_learning.html. Acesso em: 13 nov. 2022.

SCHWANDT, M. et al. Development and test of gap filling procedures for solar radiation data of the Indian SRRA measurement network. **Energy Procedia**, v. 57, p. 1100-1109, 2014.

SHI, G. et al. Data Quality Assessment and the Long-Term Trend of Ground Solar Radiation in China. **Journal of Applied Meteorology and Climatology**, v. 17, p. 1006-1016, 2008.

SILVA, M. B. P. et al. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 160, p. 11-27, 2017.

SKARVEIT, A.; OLSEN, J.; TUFT, M. E. An Hourly Diffuse Fraction Model with Correction for Variability and Surface Albedo. **Solar Energy**, v. 63, p. 171-183, 1998.

THE METEOROLOGICAL RESOURCE CENTER. **Quality Assurance and Quality Control**. 2002. Disponível em: http://www.webmet.com/met_monitoring/8.html. Acesso em: 02 set. 2019.

TREGENZA, P. R. et al. **Guide to recommended practice of daylight measurement**. Vienna: Cimmission Internationale de l'éclairage (CIE), 1994. Disponível em: <https://archive-ouverte.unige.ch/unige:103820>. Acesso em: 06 nov. 2018.

VAREJÃO-SILVA, M. A. **Meteorologia e Climatologia**. 2. ed., Recife: 2006. Disponível em: http://www.icat.ufal.br/laboratorio/clima/data/uploads/pdf/METEOROLOGIA_E_CLIMATOLOGIA_VD2_Mar_2006.pdf. Acesso em: 08 set. 2018.

VENTURA, T. M. et al. Uma abordagem computacional para o preenchimento de falhas em dados micro meteorológicos. **Revista Brasileira de Ciências Ambientais**, n. 27, p. 61-70, 2013.

VIGNOLA, F.; MICHALSKY, J.; STOFFEL, T. **Solar and Infrared Radiation Measurements**. Boca Raton: CRC Press - Taylor & Francis Group, 2012.

VOYANT, C. et al. Machine learning methods for solar radiation forecasting: A review. **Renewable Energy**, v. 105, p. 569-582, 2017.

WANG, et al. A New Outlier Detection Method Based on Optics. **Sustainable Cities and Society**, v. 45, p. 197-212, 2019.

WMO. **Guide to Meteorological Instruments and Methods of Observation**. Geneva: World Meteorological Organization, n. 8, 2012.

YANG, K.; HUANG, G. W.; TAMAI, N. A hybrid model for estimating global solar radiation. **Solar Energy**, v. 70, p. 13-22, 2001.

YAO, W.; LI, Z.; WANG, Y.; JIANG, F.; HU, L. Evaluation of global solar radiation models for Shanghai, China. **Energy Conversion and Management**, v. 84, p. 597-612, 2014.

YOUNES, S.; CLAYWELL, T.; MUNEER, T. Quality control of solar radiation data: Present status and proposed new approaches. **Energy**, v. 30, p. 1533-1549, 2005.

ZARZO, M.; MARTÍ, P. Modeling the variability of solar radiation data among weather stations by means of principal components analysis. **Applied Energy**, v. 88, p. 2775-2784, 2011.