

DETECÇÃO DE OUTLIERS DA IRRADIÂNCIA SOLAR GLOBAL POR MEIO DO Z-SCORE EM FUNÇÃO DA SAZONALIDADE E DO PADRÃO DE COBERTURA DE CÉU

Daniele Cristina Lopes Mariano - daniele.lopes@unesp.br

Marcus Vinícius Contes Calça

Matheus Rodrigues Raniero

José Rafael Franco

Carlos Gabriel dos Santos Modesto

Alexandre Dal Pai

Universidade Estadual Paulista - UNESP, Faculdade de Ciências Agronômicas, Botucatu (SP) - Brasil

Fernando de Lima Caneppele

Universidade de São Paulo - USP, Faculdade de Zootecnia e Engenharia de Alimentos, Pirassununga (SP) - Brasil

Resumo. Medir a radiação solar na superfície terrestre é mais difícil do que outras variáveis climatológicas, devido as falhas técnicas, manutenção precária e natureza dinâmica dos processos físicos, o que causa outliers no sistema de medição. O objetivo deste estudo foi detectar, quantificar e remover outliers da irradiância solar global, que são valores que apresentam um comportamento atípico, a partir de medidas coletadas a cada 1 minuto (W/m^2) do período diurno de 2022, obtidas na Faculdade de Ciências Agronômicas (UNESP) de Botucatu (SP) - Brasil. Portanto, foi utilizado o método estatístico Z-Score para calcular o quanto um valor mensurado estava distante da média populacional em relação aos demais, em termos de desvios padrão, considerando a subdivisão dos dados em função da sazonalidade e dos padrões de cobertura de céu. Neste sentido, medições de irradiância solar com Z-Score abaixo de -2 ou acima de 2 foram consideradas outliers, quantificadas e removidas do conjunto de dados. Considerando a nebulosidade, em condições de céu aberto, parcialmente aberto, parcialmente nublado e nublado, 4,47%, 2,00%, 4,00% e 4,27% do total de medidas obtidas, em cada cobertura de céu, respectivamente, foram removidas por serem outliers. Considerando a sazonalidade, no verão, primavera, inverno e outono, 3,33%, 4,42%, 3,87% e 4,48% do total de valores medidos, em cada época do ano, foram removidos por serem outliers. O Z-Score se mostrou uma técnica estatística eficiente e eficaz para detectar outliers, porém quando ocorrem muitos picos as anomalias não são identificadas em sua totalidade, pois os valores extremos afetam a média e o desvio padrão do conjunto de dados, reduzindo a capacidade de identificação de outliers. Portanto, sugere-se primeiro delimitar os valores fisicamente possíveis da irradiância solar global, obtida na superfície terrestre, no sentido de reduzir a ocorrência de valores extremos, para então identificar os outliers usando o Z-Score.

Palavras-chave: Controle de Qualidade. Detecção de Medidas Errôneas. Validação de Medidas de Energia Solar.

1. INTRODUÇÃO

A energia solar disponível em um plano imaginário no topo da atmosfera é denominada de radiação solar extraterrestre, quando atravessa a atmosfera terrestre e atinge uma superfície horizontal em uma determinada localidade, é chamada de radiação solar global/total, e pode ser particionada em duas componentes, a direta e a difusa. Neste sentido, a radiação solar direta é aquela que provém diretamente do disco solar sem sofrer os processos de absorção ou espalhamento na atmosfera terrestre, e a radiação solar difusa, é a que resulta da ação do espalhamento pelos gases, partículas e outros constituintes atmosféricos (Pereira *et al.*, 2017). Medir a radiação solar é crucial, pois permite avaliar a disponibilidade de energia solar em locais específicos do planeta Terra, contribuindo também para a análise das propriedades e distribuição dos constituintes atmosféricos, além de permitir validar informações provenientes de satélites em órbita da Terra (IQBAL, 1983; WMO, 2012). É uma ocupação que guarda grande interesse em estudos da climatologia da radiação solar, análises e avaliações de técnicas de aproveitamento da energia solar e modelagem estatística.

No entanto, medir corretamente a radiação solar disponível na superfície terrestre é mais difícil do que outras variáveis climatológicas (Calça, *et al.*, 2019). Isto se dá devido a ocorrência de falhas técnicas, de origem instrumental, de manutenção precária, de origem operacional, e da natureza dinâmica da radiação solar, pois sofre interferência da atmosfera terrestre e de outros fatores geográficos (Younes, Claywell & Muneer, 2005; Moradi, 2009). Neste sentido, torna-se necessário, a utilização de procedimentos de controle de qualidade, no intuito de detectar valores errôneos (outliers), que derivam dessas adversidades técnicas, operacionais e naturais, na intenção de não comprometer a veracidade de estudos, análises e avaliações posteriores do recurso de energia solar disponível. Portanto, o objetivo deste estudo foi criar um procedimento de detecção de outliers da irradiância solar global coletada na superfície terrestre, com o intuito de identificar valores que apresentam um comportamento atípico, a partir de medidas instantâneas (1 minuto em W/m^2) do ano de 2022, obtidas na Faculdade de Ciências Agronômicas (UNESP) de Botucatu (São Paulo) - Brasil.

2. MATERIAL E MÉTODOS

2.1 Local de Estudo

Este estudo foi elaborado utilizando as medidas de irradiância solar global obtidas pelo 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, São Paulo - Brasil. Situado na região centro-sul do estado, com 149 mil habitantes e uma área territorial de 1.482,642 km², Botucatu - SP (Fig. 1) possui um bioma caracterizado por uma vegetação de Cerrado e Mata Atlântica (IBGE, 2022), com fortes atividades agrícolas, que compreendem grandes plantações de cana-de-açúcar e eucalipto, assim como atividades industriais e comerciais moderadas (Codato *et al.*, 2008; Dal Pai *et al.*, 2016). Sua formação geográfica compreende um relevo de colinas não simétricas 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, que é a mais alta. Esta formação geográfica provoca diferenças climáticas consideráveis no município (Silva *et al.*, 2017).

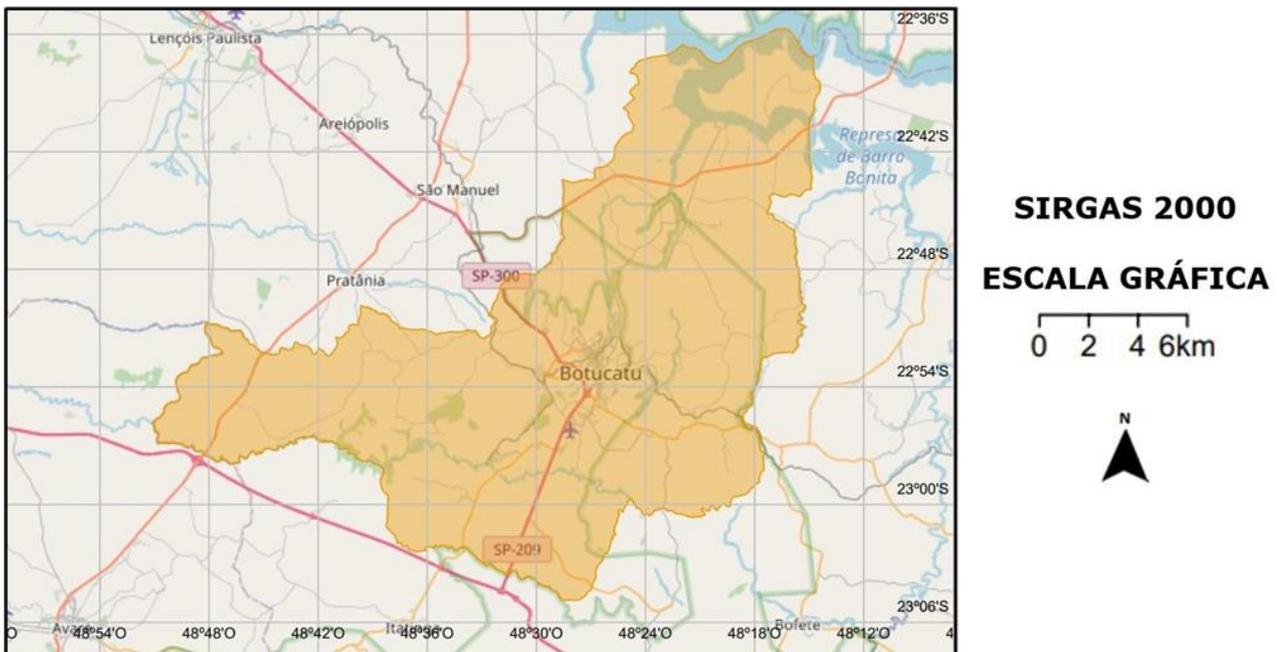


Figura 1 - Localização geográfica do Município de Botucatu (São Paulo) - Brasil.

O clima de Botucatu - SP é identificado como Aw, de acordo com a classificação climática de Köppen (Franco, *et al.*, 2023), compreendendo um inverno seco (junho a agosto) e um verão quente e chuvoso (dezembro a fevereiro). De acordo com a série histórica de 1991 a 2020 da Estação Meteorológica do Departamento de Engenharia Rural e Socioeconomia da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu (SP):

- Fevereiro é o mês mais quente, com temperatura média do ar de 23,80 °C.
- Agosto é o mês mais frio, com temperatura média do ar de 18,28 °C.
- Janeiro é o mês mais úmido, com umidade relativa média do ar de 75,76%.
- Agosto é o mês menos úmido, com umidade relativa média do ar de 61,87%.
- Janeiro é o mês mais chuvoso, com uma média pluviométrica de 315mm.
- Agosto é o mês menos chuvoso, com uma média pluviométrica de 38mm.

2.2 Instrumentação e Medidas

A irradiância solar global na superfície terrestre foi medida por um Piranômetro da Eppley - PSP, em um plano horizontal não sombreado, instalado em uma área rural com grama verde e curta a 1,5 metros do solo (Fig. 2). Um Micrologger CR3000 da Campbell Scientific foi utilizado para coletar, pré-processar e gravar no local as medidas a cada 1 minuto (média das observações obtidas a cada 5 segundos) no ano de 2020. Após a coleta, as medições eram transmitidas para um servidor de dados a partir de um módulo de radiofrequência da Campbell Scientific - RF450, onde eram estocadas no Sistema de Gerenciamento de Banco de Dados Relacional - MySQL, para posterior uso. Uma rotina regular de manutenção foi realizada, no intuito de fornecer as condições ideais de monitoramento da irradiância solar, de acordo com as diretrizes da Organização Mundial de Meteorologia (WMO, 2012). O Piranômetro da Eppley - PSP, usado para obter a irradiância solar global, foi calibrado no final do ano de 2021, a partir de outro Piranômetro Eppley - PSP de referência, acondicionado de forma íntegra em um ambiente protegido e utilizado somente para esta finalidade.



Figura 2 - Instrumentos da Estação de Radiometria Solar da Faculdade de Ciências Agrônômicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu (SP) - Brasil.

2.3 Procedimento de Detecção de Outliers

As medidas de irradiância solar obtidas na superfície terrestre tem cada vez mais se popularizado, especialmente pelo efeito das mudanças climáticas globais e sua relação com a energia solar. Porém, é importante levar em consideração que uma medida de irradiância solar pode ter sua consistência afetada por adversidades instrumentais (imprecisões nos instrumentos de medição) ou por negligências operacionais (Tab. 1) no local onde são dispostos os instrumentos de monitoramento da irradiância solar disponível na superfície terrestre (Muneer & Fairouz, 2002).

Tabela 1 - Principais fontes de erros em medidas de irradiância solar.

ADVERSIDADES INSTRUMENTAIS	NEGLIGÊNCIAS OPERACIONAIS
Resposta do cosseno	Excremento de aves
Resposta da temperatura	Orvalho na cúpula dos sensores
Seletividade espectral	Nivelamento incorreto dos sensores
Desalinhamento do anel de sombreamento	Mal funcionamento do cabeamento
Estabilidade do instrumento	Quedas na eletricidade

Fonte: Muneer, Younes & Munawwar (2007).

A avaliação de medidas de irradiância solar, para a detecção de outliers e validação dos valores obtidos, é um procedimento que deve ser realizado a partir de diferentes princípios (Ohmura *et al.*, 1998):

- **Princípio Físico:**
Tem como objetivo de identificar se os valores medidos estão dentro dos limites fisicamente possíveis.
- **Princípio Instrumental:**
Tem como objetivo de analisar a diferença dos valores medidos por sensores diferentes no mesmo local.
- **Princípio Estatístico:**
Tem como objetivo de identificar valores que se destacam significativamente em relação aos demais.

É importante também considerar o comportamento sazonal das medidas de irradiância solar, de acordo com as estações do ano e a natureza dinâmica da radiação solar (influência do local geográfico e dos constituintes atmosféricos).

Neste sentido, no intuito de fornecer uma maior segurança para estudos que envolvem o uso da irradiância solar global, visando a utilização de medidas com valores mais precisos, pode-se aplicar métricas estatísticas de detecção de outliers. Um outlier é um ponto de dados que se distancia notavelmente do padrão geral ou da tendência observada em um conjunto de medidas de irradiância solar. Identificar e lidar com eles é necessário, pois podem distorcer a interpretação dos resultados de investigações científicas e afetar a precisão de modelos estatísticos. O Z-Score é um método comum,

mas eficiente, para a detecção de outliers, seu objetivo é descrever o quanto um valor está distante da média populacional em termos de desvios padrão (Curtis *et al.*, 2016). Portanto a utilização do Z-Score (Eq. 1) torna possível estabelecer critérios para a detecção de valores atípicos em medidas de irradiância solar, tornando os dados padronizados e amplamente compreensíveis em relação a distribuição geral do período considerado, além disso, facilita a tomada de decisão, pois estabelece critérios numéricos que permitem decidir remover ou manter a informação no conjunto de dados.

$$Z = \frac{(X - \mu)}{\sigma} \tag{1}$$

Na Eq. 1, Z representa a pontuação Z (Z-Score), X é a medida de irradiância solar global que está sendo avaliada, μ é a média da irradiância solar global do conjunto de dados total e σ é o desvio padrão da irradiância solar global do conjunto de dados total. Um Z-Score acima da média populacional terá um valor positivo, enquanto um Z-Score abaixo da média terá seu valor será negativo, portanto, quanto maior o desvio do Z-Score, em valores positivos ou negativos, maior será a magnitude do desvio da média (Curtis *et al.*, 2016).

A curva de distribuição normal (Fig. 3) é comumente utilizada para descrever a frequência em que valores ocorrem em um histograma, de modo que a média é posicionada no meio da curva, enquanto o desvio padrão fornece o grau de dispersão dos dados. Para cada subdivisão da curva normal há uma porcentagem de dados correspondentes, sendo que aproximadamente 68,26% da área da curva está a um desvio padrão da média, cerca de 95,44% estão a dois desvios padrão da média e por fim, 99,72% se encontram a três desvios padrão da média (Caire, 2013). A aplicação da curva de distribuição normal em conjuntos de dados avaliados utilizando o Z-Score viabiliza a definição do limiar de filtragem das medidas de irradiância solar consideradas outliers, pois, conforme a quantidade de desvios padrão empregados, a proporção de valores identificados como discrepantes é reduzida. Neste estudo, considerou-se a adoção de dois desvios padrão do limite de confiança, que sinaliza uma medida considerada correta de irradiância solar global, possibilitando, então, a remoção de valores que se situavam fora do intervalo de confiança especificado.

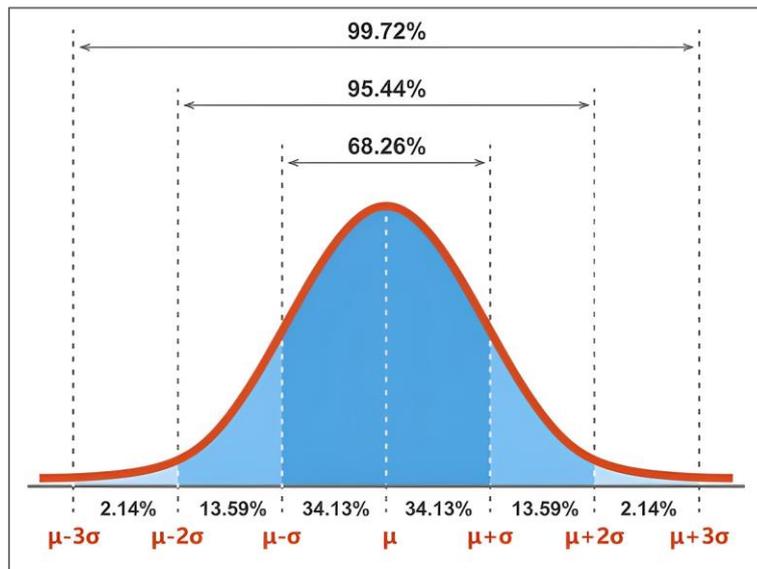


Figura 3 - Diagrama esquemático da distribuição normal gaussiana.

Fonte: Adaptado de Caire (2013).

Conforme citado anteriormente, uma das dificuldades que se tem em medir a irradiância solar é o seu comportamento sazonal, de acordo com as estações do ano, e a natureza dinâmica da intensidade dos valores captados, devido a influência do local geográfico e dos constituintes atmosféricos no momento que a medida é obtida. Neste sentido, a fim de melhorar a detecção de outliers, foi realizada a classificação sazonal do conjunto de dados, sinalizando as épocas do ano em função do dia juliano (Tab. 2), que representa os dias decorridos ao longo do ano.

Tabela 2 - Classificação da sazonalidade em função do dia juliano.

CLASSIFICAÇÃO DA SAZONALIDADE	DIAS JULIANO (DJ)
Verão	01 ao 79 e do 355 ao 365
Outono	80 ao 171
Inverno	172 ao 265
Primavera	266 ao 354

No sentido de considerar, também, a intensidade dos valores de irradiância solar de acordo com os constituintes atmosféricos quando a medida é obtida, foi feita uma segunda classificação, dos padrões de cobertura de céu para cada valor medido (Tab. 3). A classificação da cobertura de céu foi realizada em função do índice de transmissividade atmosférica (KT), que representa a razão (G/E) entre a irradiância solar global medida (G) e a irradiância solar extraterrestre calculada (E) na escala temporal instantânea (1 minuto) em W/m², de acordo com Escobedo *et al.* (2009).

Tabela 3 - Classificação da cobertura de céu em função do índice de transmissividade atmosférica (KT).

PADRÃO DE COBERTURA DE CÉU	ÍNDICE DE TRANSMISSIVIDADE (KT)
Nublado	$0 \leq KT < 0,35$
Parcialmente Nublado	$0,35 \leq KT < 0,55$
Parcialmente Aberto	$0,55 \leq KT < 0,65$
Aberto	$0,65 \leq KT < 1$

Após os dados serem identificados de acordo com a classificação da sazonalidade e o padrão de cobertura de céu, foi possível então determinar quantos valores medidos de irradiância solar global pertenciam a cada grupo, e de acordo com essa divisão, iniciar as análises estatísticas, considerando então as épocas do ano e a variação causada pelos constituintes atmosféricos no momento que a medida é obtida. Na Tab. 4, se tem, portanto, a quantidade total de medidas de irradiância solar global disponível para cada época do ano e seus padrões de cobertura de céu em 2022.

Tabela 4 - Quantidade total disponível de medidas de irradiância solar global em 2022 de acordo com a sazonalidade e o padrão de cobertura de céu.

SAZONALIDADE	COBERTURA DE CÉU	QUANTIDADE DE VALORES
Verão	Nublado	26.715
	Parcialmente Nublado	10.821
	Parcialmente Aberto	5.055
	Aberto	20.112
Outono	Nublado	17.609
	Parcialmente Nublado	10.740
	Parcialmente Aberto	7.665
	Aberto	24.627
Inverno	Nublado	15.492
	Parcialmente Nublado	11.899
	Parcialmente Aberto	8.765
	Aberto	25.012
Primavera	Nublado	20.566
	Parcialmente Nublado	8.888
	Parcialmente Aberto	4.958
	Aberto	20.619

O procedimento de detecção de outliers nas medidas de irradiância solar global instantâneas (1 minuto) em W/m² foi criado por meio da linguagem de programação R. Portanto, as medidas foram exportadas do Sistema de Gerenciamento de Banco de Dados Relacional - MySQL (conforme descrito na Seção 2.2), onde se encontram estocadas a longo prazo, e foram carregadas no ambiente de análise de dados do R Studio - Versão 4.2.2, onde foi possível executar os cálculos para obter o Z-Score e utilizar a biblioteca GGPlot 2 para as análises gráficas (histograma e gráficos de distribuição).

3. RESULTADOS E DISCUSSÃO

Desde o início de 2019 o Laboratório de Agrometeorologia e Radiometria Solar da Faculdade de Ciências Agrônomicas da Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP de Botucatu, São Paulo - Brasil coleta medidas de irradiância solar, na superfície terrestre, a cada 1 minuto do dia. Obter informações de energia solar com essa frequência de aquisição de dados fornece a capacidade de capturar variações rápidas e sutis na intensidade da luz solar, porém, em contrapartida, exige uma atenção maior para garantir a precisão e confiabilidade das medidas coletadas, principalmente considerando as condições instrumentais e ambientais adversas que podem ocorrer, também, com mais frequência. Portanto, utilizando as medidas de irradiância solar global adquiridas no ano de 2022, classificadas de acordo com a sazonalidade e o padrão de cobertura de céu, foi gerado o Z-Score para cada uma delas, permitindo

identificar os valores discrepantes de forma mais fácil. Neste caso, foi feita a plotagem de histogramas de frequência do Z-Score, separados por cada estação do ano e de acordo com cada padrão de cobertura de céu (Fig. 4).

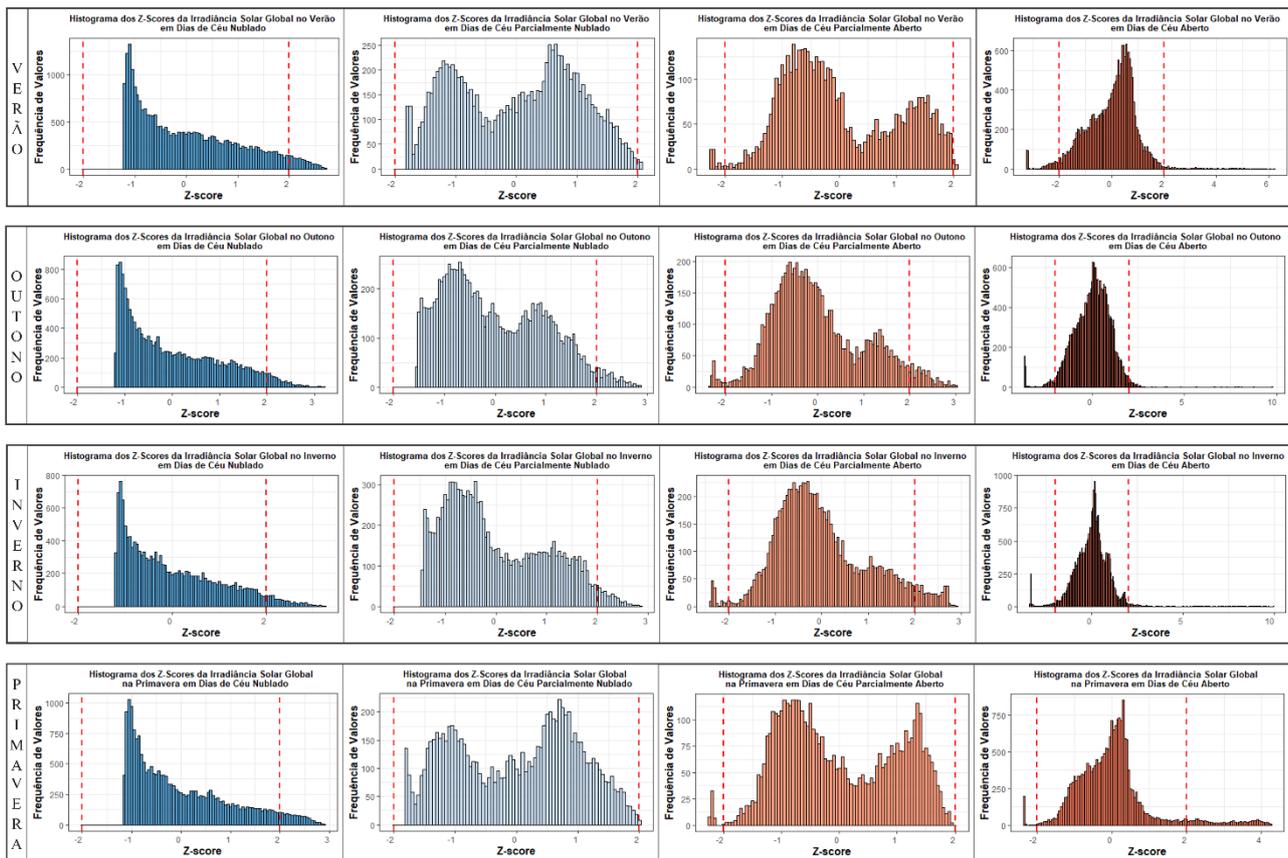


Figura 4 - Histogramas dos Z-Scores da irradiância solar global obtidos de acordo com a sazonalidade e o padrão de cobertura de céu.

É possível observar que os histogramas dos Z-Scores apresentam uma distribuição da frequência de ocorrência dos valores semelhante em todas as estações do ano, quando levado em consideração o padrão de cobertura de céu. Além disso, a decisão de realizar a classificação e a segmentação das medidas de irradiância solar, para posteriormente executar o cálculo do Z-Score, em função a sazonalidade e do padrão de cobertura de céu, trouxe melhores parâmetros estatísticos para a identificação de outliers, uma vez que se leva em consideração a intensidade da luz solar que atinge a superfície terrestre de acordo com a sua interação com os constituintes atmosféricos. Se os valores medidos não estivessem classificados adequadamente, no sentido de obter o Z-Score para todo o conjunto de dados disponibilizado, as médias e os desvios padrões seriam calculados de forma mais abrangente, sem considerar os efeitos atmosféricos e astronômicos que normalmente ocorrem com a irradiância solar, podendo dar a entender que um certo grupo de valores localizados nas extremidades dos histogramas fossem considerados confiáveis.

As linhas verticais tracejadas na cor vermelha têm como finalidade indicar os limites estabelecidos para considerar uma medida de irradiância solar confiável, pois, neste caso, ocorrências abaixo de -2 ou acima de 2 desvios padrão indicam que um valor obtido é um outlier, uma vez que se distancia consideravelmente do padrão de variação dos demais valores. No verão e na primavera foi observado que os valores medidos mais discrepantes, em relação à média geral, foram obtidos em condições de céu nublado e aberto, respectivamente. Neste sentido, na condição de céu nublado não houveram Z-Scores negativos, já em condições de céu aberto obteve-se Z-Scores dos dois lados da distribuição das ocorrências (negativo e positivo). Em relação as condições de céu parcialmente nublado e parcialmente aberto, a ocorrência de Z-Scores fora do intervalo de dois desvios padrão foram menores, onde especificamente na condição de céu parcialmente nublado não houveram Z-Scores negativos, e na condição de céu parcialmente aberto ocorreram negativos e positivos. No verão a frequência dos picos de Z-Score está acima de 1500 ocorrências, já na primavera chega próximo a 1125 ocorrências.

Para os períodos de outono e inverno os Z-Scores obtidos apresentaram semelhanças com as sazonalidades anteriores. Neste caso, para o outono o histograma que apresentou os maiores picos do Z-Score foi o da condição de céu nublado, com pouco mais de 800 ocorrências, sendo que se encontravam a menos um desvio padrão da média geral. Para o inverno a maior frequência de picos do Z-Score foi identificada na condição de céu aberto, com aproximadamente 900 ocorrências, sendo que estavam mais próximas da média geral dos valores.

Antes de determinar a remoção dos valores identificados como outliers a partir do Z-Score, levando em consideração as variações naturais da irradiância solar ao longo do ano e as condições atmosféricas no momento de sua obtenção, foi

realizada uma análise gráfica dos valores coletados com o auxílio de gráficos de dispersão em relação ao tempo. Notou-se que, todas as estações do ano continham medidas com um comportamento fora do padrão esperado, que é mais um indicativo de serem outliers. Em condições normais, a irradiância solar disponível na superfície terrestre é menor do que a quantidade total de energia solar que chega ao topo da atmosfera, devido a sua atenuação pelos fenômenos físicos de reflexão e espalhamento causados pelos constituintes atmosféricos. Neste sentido, um comportamento padrão é a intensidade de irradiância solar medida na superfície terrestre, a partir de piranômetros, ser menor que a irradiância solar que atinge o topo da atmosfera, que neste caso, tem o seu valor máximo para cada estação do ano representado pela cor vermelha nos gráficos de dispersão apresentados na Fig. 5. Portanto, para o verão (Fig. 5a), foram identificadas medidas de irradiância solar obtidas na superfície terrestre com seus valores máximos variando de 1500 a 2500 W/m², no outono (Fig. 5b) identificou-se a variação dos valores máximos ocorrendo de 1000 a 2500 W/m², no inverno (Fig. 5c) notou-se também as variações dos valores máximos em uma escala de 1000 a 2500 W/m², e por fim, na primavera (Fig. 5d) os valores máximos variaram de 1500 a 2500 W/m². Ou seja, todos os pontos de dispersão acima da linha vermelha indicam uma medida fisicamente impossível.

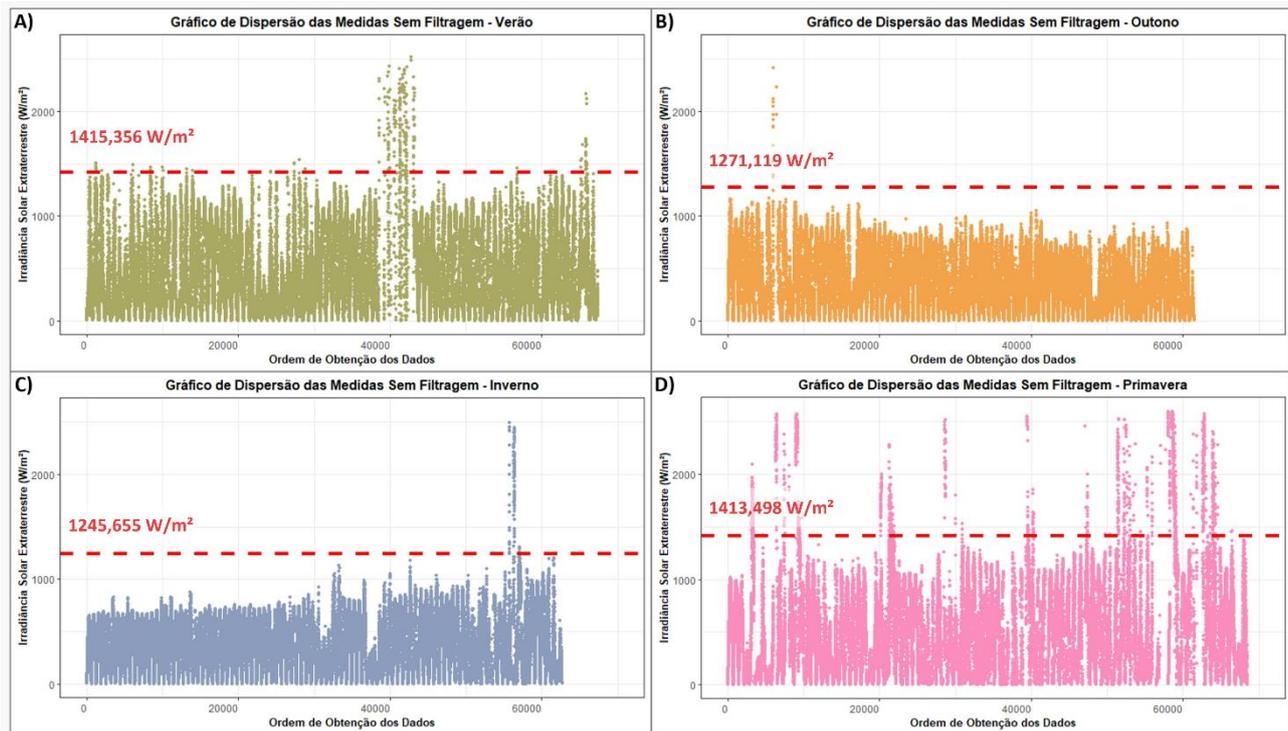


Figura 5 - Gráfico de dispersão das medidas de acordo com a sazonalidades antes da remoção dos outliers.

Após a identificação dos pontos de dispersão que representam as medidas de irradiância solar com valores fora do padrão esperado de variação (Fig. 5), considerando as estações do ano, e a definição de um ponto de corte nos histogramas do Z-Score, em relação a distribuição das ocorrências em função da sazonalidade e do padrão de cobertura de céu (Fig. 4), foram removidos os outliers detectados. Neste sentido, portanto, todas as medidas de irradiância solar na superfície terrestre que obtiveram um Z-Score abaixo de -2 e acima de 2 foram removidas do conjunto de dados.

Após a remoção dos outliers, no verão (Fig. 6a), os valores máximos das medidas da irradiância solar obtidas na superfície terrestre não ultrapassaram os 1415,356 W/m², que representa o valor máximo calculado para o topo da atmosfera. Em relação ao outono (Fig. 6b), os valores máximos já não ultrapassam mais os 1271,119 W/m², que também representa o máximo teoricamente possível. Já para o inverno (Fig. 6c) foi possível identificar uma pequena oscilação dos valores máximos mensurados na superfície da Terra, no início do período variam próximos a 750 W/m² e no final ficam mais próximos de 1000 W/m², porém, mesmo considerando essa oscilação ficam abaixo do limite máximo fisicamente possível de 1245,655 W/m² para a época do ano. Por fim, na primavera (Fig. 6d) os valores medidos de irradiância solar na superfície terrestre possuem uma oscilação mais considerável, atingindo inclusive o limite sazonal máximo calculado de 1413,498 W/m², onde, devido a grande quantidade de picos ocorridos antes da detecção de outliers (Fig. 5d), o valor médio e o desvio padrão foram afetados, influenciando também o cálculo do Z-Score.

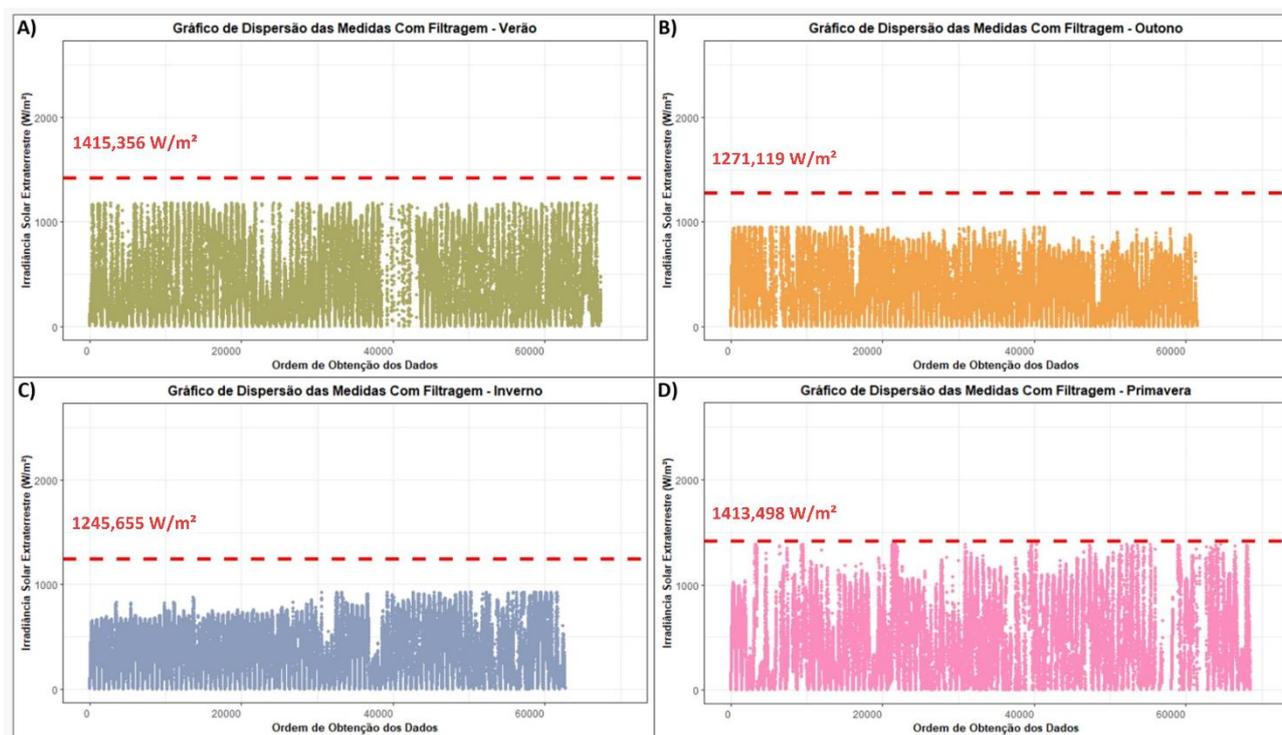


Figura 6 - Gráfico de dispersão das medidas de acordo com a sazonalidade após a remoção dos outliers.

Inicialmente, foram selecionadas 239.543 medidas de irradiância solar global, referentes somente ao período diurno, no intuito de executar o procedimento de detecção de outliers, sendo que, apenas 9.172 valores foram classificados como outliers utilizando o Z-Score, o que representa 3,83% do total de informações disponibilizadas. Considerando o padrão de nebulosidade pôde-se obter as seguintes informações:

- Do total de 80.382 medições obtidas em céu nublado, 3.597 foram consideradas outliers (4,47%).
- Do total de 32.682 medições obtidas em céu parcialmente nublado, 655 foram consideradas outliers (2,00%).
- Do total de 26.443 medições obtidas em céu parcialmente aberto, 1.057 foram consideradas outliers (4,00%).
- Do total de 90.370 medições obtidas em céu aberto, 3.863 foram consideradas outliers (4,27%).

Considerando, além do padrão de cobertura de céu, a sazonalidade (Tab. 5), foi possível constatar que, no verão o maior e o menor percentual de outliers, em relação ao número total de valores disponíveis, foram detectados nas condições de céu aberto (4,29%) e parcialmente nublado (0,23%), respectivamente. No outono o maior e o menor percentual de outliers, em relação a quantidade total de medições disponibilizadas, foram identificados nas condições de céu parcialmente aberto (5,41%) e parcialmente nublado (2,83%), respectivamente. No inverno, o maior e o menor percentual de outliers, em relação ao número total de medidas disponíveis, foram detectados nas condições de céu parcialmente aberto (6,01%) e parcialmente nublado (2,73%), respectivamente. E, por fim, na primavera, o maior e o menor percentual de outliers, em relação a quantidade total de medições disponibilizadas, foram identificados nas condições de céu aberto (6,36%) e parcialmente nublado (0,01%), respectivamente. Levando em consideração apenas a sazonalidade, no outono foi detectado o maior percentual de outliers (4,48%), seguido da primavera (4,42%), do inverno (3,87%) e, por último, do verão (3,33%), em relação ao total de medidas da irradiância solar global, obtidas em cada estação do ano, para o período em questão.

Tabela 5 - Quantificação de outliers em relação aos totais de medidas disponíveis de irradiância solar global.

SAZONALIDADE	COBERTURA DE CÉU	TOTAL DE VALORES	QUANTIDADE DE OUTLIERS	PERCENTUAL DE OUTLIERS
Verão	Nublado	26.715	1.134	4,24%
	Parcialmente Nublado	10.821	25	0,23%
	Parcialmente Aberto	5.055	67	1,33%
	Aberto	20.112	862	4,29%
Outono	Nublado	17.609	705	4,00%
	Parcialmente Nublado	10.740	304	2,83%
	Parcialmente Aberto	7.665	415	5,41%
	Aberto	24.627	861	3,50%
Inverno	Nublado	15.492	681	4,40%

	Parcialmente Nublado	11.899	325	2,73%
	Parcialmente Aberto	8.765	527	6,01%
	Aberto	25.012	836	3,34%
Primavera	Nublado	20.566	1.077	5,24%
	Parcialmente Nublado	8.888	1	0,01%
	Parcialmente Aberto	4.958	48	0,97%
	Aberto	20.619	1.304	6,32%
TOTAL		239.543	9.172	3,83%

Apesar dos outliers representarem cerca de 3,83% (9.172 valores) do total das medidas de irradiância solar global (239.543 valores), consideradas neste estudo, é um valor baixo, levando em conta que a coleta de medidas ocorre a cada 1 minuto durante o período diurno, obtendo uma média aproximada de 780 valores obtidos para o verão, 670 para a primavera, 665 para o inverno e 770 para o outono, em Botucatu (São Paulo) - Brasil. Neste sentido, convertendo o percentual de outliers detectados em uma quantidade equivalente a dias de medições da irradiância solar global removidos, pode-se considerar uma quantidade total correspondente a aproximadamente 13 dias de valores mensurados. Porém, é importante ponderar também que, na maior parte das ocasiões, os outliers que foram removidos do conjunto de dados não ocorreram para um mesmo período do dia, nem mesmo para um mesmo dia, mas espaçadamente durante a época do ano em que foram investigados. Portanto, ao realizar a integração numérica dos valores de irradiância solar global obtidos a cada 1 minuto para o período de 1 dia, é possível que não haja influência considerável no valor final.

4. CONCLUSÃO

Neste estudo, pôde-se obter as seguintes conclusões:

Executar o procedimento de detecção de outliers da irradiância solar global medida a cada 1 minuto em Botucatu (São Paulo) - Brasil, por meio do Z-Score calculado para os valores obtidos e classificados de acordo com a sazonalidade e o padrão de cobertura de céu, permitiu uma melhor adequação dos parâmetros estatísticos do conjunto de dados, uma vez que foi levado em consideração os níveis da irradiância solar disponível na superfície terrestre de acordo com os fatores astronômicos e a sua interação com os constituintes atmosféricos.

O Z-Score é uma técnica estatística que permite definir um intervalo comum de variação para as medidas de irradiância solar, em função da média e do desvio padrão dos valores de um conjunto de dados, sem a necessidade de usar parâmetros externos para avaliar a confiabilidade do que foi mensurado. Neste sentido, em um contexto geral 3,83% do total de medidas de irradiância solar, coletadas na superfície terrestre em 2022, foram removidas do conjunto de dados por terem um Z-Score abaixo de -2 ou acima de 2, considerando a separação das informações em função da sazonalidade e do padrão de cobertura de céu. Portanto, a maior e a menor ocorrência de outliers, considerando o padrão de cobertura de céu, ocorreu em condições nubladas (4,47% do total) e parcialmente nubladas (2,00% do total), respectivamente. De modo que, considerando a sazonalidade, a maior e a menor ocorrência de outliers foram para o outono (4,48% do total) e verão (3,33%), respectivamente. Observou-se que os valores não removidos se encontravam abaixo do limite máximo teoricamente aceito.

Apesar do Z-Score ter se mostrado uma técnica estatística eficiente e eficaz para detectar outliers, quando ocorrem muitos picos nos valores as anomalias não são identificadas totalmente, pois os valores extremos afetam a média e o desvio padrão do conjunto de dados, reduzindo a capacidade de se identificar outliers usando o Z-Score, como foi observado para a primavera. Portanto, sugere-se utilizar primeiro métodos que delimitem os valores fisicamente possíveis que podem ocorrer, para depois usar o Z-Score para identificar os outliers.

Agradecimentos

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Faculdade de Ciências Agrárias (UNESP) de Botucatu (São Paulo) - Brasil pelo apoio financeiro concedido através das bolsas de estudos (Mestrado e Doutorado).

REFERÊNCIAS

- Calça, M. V. C.; Raniero, M. R.; Fernando, D. M. Z.; Rodrigues, S. A.; Dal Pai, A., 2019. Outliers Detection in a Quality Control Procedure for Measurements of Solar Radiation. IEEE Latin America Transactions, vol. 17, n. 11, pp. 1815-1822.
- Codato, G.; Oliveira, A. P.; Soares, J.; Escobedo, J. S.; Gomes, E. N.; Dal Pai, A., 2008. Global and diffuse solar irradiances in urban and rural areas in southeast Brazil. Theoretical and Applied Climatology, vol. 93, pp. 57-73.
- Curtis, A. E.; Smith, T. A.; Zynganshin, B. A.; Eleftheriades, J. A., 2016. The Mystery of the Z-Score. Aorta, vol. 4, n. 4, pp. 124-130.

- Dal Pai, A.; Escobedo, J. F.; Dal Pai, E.; Oliveira, A. P.; Soares, J.; Codato, G., 2016. MEO shadowring method for measuring diffuse solar irradiance: Corrections based on sky cover. *Renewable Energy*, vol. 99, pp. 754-763.
- Escobedo, J. F.; Gomes, E. N.; Oliveira, A. P.; Soares, J., 2009. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. *Applied Energy*, vol. 86, pp. 299-309.
- Franco, J. R.; Dal Pai, E.; Calça, M. V. C.; Raniero, M. R.; Dal Pai, A.; Sarnighausen, V. C. R.; Román, R. M. S., 2023. Atualização da normal climatológica e classificação climática de Köppen para o município de Botucatu - SP. *Irriga - Brazilian Journal of Irrigation and Drainage*, vol. 28, n. 1, pp. 77-92.
- Iqbal, M., 1983. *An Introduction to Solar Radiation*. Ontario: Academic Press Canada.
- Moradi, I., 2009. Quality control of global solar radiation using sunshine durations hours. *Energy*, vol. 34, pp. 1-6.
- Muneer, T.; Fairouz, F., 2002. Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases. *Building Services Engineering Research and Technology*, vol. 23, pp. 151-166.
- Muneer, T.; Younes, S.; Munawwar., 2007. Discourses on solar radiation modeling. *Renewable and Sustainable Energy*, vol. 11, pp. 551-602.
- Ohmura, A.; Dutton, E. G.; Forgan, B.; Fröhlich, C.; Gilgen, H.; Hegner, H.; Heimo, A.; König-Langlo, G.; McArthur, B.; Müller, G.; Philipona, R.; Pinker, R.; Whitlock, C. H.; Wild, M., 1998. Baseline Surface Radiation Network (BSRN/WCRPP): New Precision Radiometry for Climate Research. *Bulletin of the American Meteorological Society*, vol. 79, pp. 2115-2136.
- Pereira, E. B.; Martins, F. R.; Gonçalves, A. R.; Costa, R. S.; Lima, F. L.; Rüther, R.; Abreu, S. L.; Tiepolo, G. M.; Pereira, S. V.; Souza, J. G., 2017. Atlas Brasileiro de Energia Solar. 2 ed., São José dos Campos: Instituto Nacional de Pesquisas Espaciais. Disponível em: <http://doi.org/10.34024/978851700089>.
- Silva, M. B. P.; Escobedo, J. F.; Santos, C. M.; Rossi, T. J.; Silva, S. H. M. G., 2017. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 160, pp. 11-27.
- WMO., 2012. *Guide to Meteorological Instruments and Methods of Observation*. Geneva: World Meteorological Organization, n. 8.
- Younes, S.; Claywell, T.; Muneer, T., 2005. Quality control of solar radiation data: Present status and proposed new approaches. *Energy*, vol. 30, pp. 1533-1549.

GLOBAL SOLAR IRRADIANCE OUTLIERS DETECTION USING Z-SCORE BASED ON SEASONALITY AND SKY COVERAGE

Abstract. *Measuring solar radiation on the Earth's surface is more difficult than other climatological variables, due to technical failures, poor maintenance, and the dynamic nature of physical processes, which causes outliers in the measurement system. The aim of this study was to detect, quantify and remove outliers in global solar irradiance, which are values that present atypical behavior, based on measurements collected every 1 minute (W/m^2) of the 2022 daytime period, obtained at the School of Agricultural Sciences (UNESP) of Botucatu (SP) - Brazil. Therefore, the Z-Score statistical method was used to calculate how far a measured value was from the population average in relation to the others, in terms of standard deviations, considering the subdivision of the data depending on seasonality and sky coverage patterns. In this sense, solar irradiance measurements with Z-Score below -2 or above 2 were considered outliers, quantified, and removed from the data set. Considering cloudiness, in clear, partially clear, partially cloudy, and cloudy conditions, 4.47%, 2.00%, 4.00% and 4.27% of the total measurements obtained, in each sky coverage, respectively, were removed because they were outliers. Considering seasonality, in summer, spring, winter and autumn, 3.33%, 4.42%, 3.87% and 4.48% of the total values measured, at each time of the year, were removed because they were outliers. Z-Score proved to be an efficient and effective statistical technique for detecting outliers, however, when many peaks occur, anomalies are not identified in their entirety, as extreme values affect the mean and standard deviation of the data set, reducing the ability to identification of outliers. Therefore, it is suggested to first delimit the physically possible values of global solar irradiance, obtained on the Earth's surface, to reduce the occurrence of extreme values, and then identify outliers using the Z-Score.*

Keywords: *Quality Control. Erroneous Measurements Detection. Solar Energy Measurements Validation.*