

PROCEDIMIENTO OBJETIVO PARA A GARANTIA DE QUALIDADE DE DADOS DE RADIAÇÃO SOLAR

L. Petribú^{1,2}, E. Sabino^{1,2,3}, H. Barros^{1,2}, A. Costa^{1,2}, E. Barbosa^{1,2}, O.C. Vilela^{1,2}

¹ Departamento de Energia Nuclear da Universidade Federal de Pernambuco (DEN-UFPE). Av. Prof. Luiz Freire, s/n, Recife-PE, Brasil.

² Centro de Energias Renováveis da Universidade Federal de Pernambuco (CER-UFPE). Av. da Arquitetura, s/n, Recife-PE, Brasil.

³Instituto de Educação, Ciência e Tecnologia de Pernambuco (IFPE), Campus Pesqueira, BR 232, km 208, Prado, Pesqueira-PE, Brasil.
e-mail: leonardo.brennand@ufpe.br

Recibido 29/07/17, aceptado 28/09/17

RESUMEN: Se propone un procedimiento objetivo para la garantía de calidad de los datos de irradiancia solar disponibles. El trabajo establece la unión de procedimientos ya difundidos en la literatura (filtros físicos y estadísticos) - juntamente con algunos ajustes propuestos - y la utilización de procedimientos inéditos (e.g. box whiskers o diagramas de caja) en lo que se refiere a la garantía de calidad de datos solarimétricos. Para la obtención de los resultados, fue aplicado el procedimiento a datos de la red SONDA, en las estaciones Petrolina - PE y Brasilia-DF. Las dos series temporales tuvieron menos de 10% de datos rechazados de irradiancia solar en sus componentes global, directa y difusa.

Palavras-chave: garantía de calidad; radiación solar; pruebas globales; diagramas de caja.

INTRODUÇÃO

A demanda por energia elétrica tem aumentado significativamente nos últimos anos, promovendo uma consolidação no cenário energético mundial das fontes renováveis, como a solar e eólica. No planejamento de centrais fotovoltaicas e heliotérmicas, um procedimento importante consiste na medição do recurso solar com acurácia para determinar a melhor localização e prever a produção de energia, dando devido suporte a questões de financiamento e licenciamento. Além disso, a utilização de dados de radiação solar se mostra bastante difundida atualmente em várias áreas de pesquisa. Contudo, os dados obtidos na campanha de medição e na operação quase sempre contêm espúrios que comprometem a qualidade e a confiabilidade da série temporal. Diversos procedimentos de garantia de qualidade de dados observacionais têm sido propostos ao longo de várias décadas (Maxwell, 1993; Younes et al., 2005; Journee e Bertrand, 2011; Raichijk, 2012).

O estudo da garantia de qualidade para tratamento de dados meteorológicos se iniciou na década de 50, quando Gilchrist e Cressman (1954) propuseram uma análise subjetiva para dados de pressão atmosférica e sugeriram estudos futuros para o emprego de análises objetivas. No enfoque da radiação solar, filtros de natureza física e estatística foram propostos por diversos autores. Maxwell et al. (1993) propuseram um filtro físico que exclui os períodos de início e final do dia solar. Os autores também utilizaram um filtro estatístico estabelecendo curvas que caracterizam limites superior e inferior (envelope), aplicados à relação entre o índice de claridade (k_t) e a transmitância direta da atmosfera (k_n). Os limites foram ajustados segundo o duplo exponencial de Gompertz. Essa metodologia também foi utilizada por Raichijk (2012), que empregou limites calculados segundo a média e o desvio padrão dos parâmetros da dispersão, assim como foram sugeridos em Younes et al. (2005). Estes últimos detalharam a causa e origem de vários erros de medição, destacando-se o erro de resposta cossenóide, causada pela incerteza do sensor de medição para baixos ângulos de incidência de radiação, e o dark-

offset, que é um erro de radiação de onda longa à noite devido a desequilíbrios térmicos, causando o retorno de valores negativos para a irradiância solar. Os autores também estabeleceram limites de aceitação para os valores de k_t e k_d (razão difusa) e testes de consistência entre parâmetros. Testes de mesma natureza também foram utilizados pela BSRN – *Baseline Surface Radiation Network*. Long e Shi (2006) propuseram um teste que verifica se o tracker está seguindo corretamente o sol. Journee e Bertrand (2011) empregaram limites utilizando a radiação solar em suas três componentes (global, direta e difusa) e as respectivas radiações de céu claro, calculadas pelo modelo ESRA - *European Solar Radiation Atlas*. Sönmez (2013) propôs testes que avaliam a repetição sucessiva de determinado valor (persistência) e saltos bruscos instantâneos na irradiância solar (step test).

Com vistas a automatizar o procedimento de garantia de qualidade de dados de irradiância solar, propõe-se neste trabalho utilizar procedimentos já estabelecidos na literatura da área de solar (com algumas sugestões de modificações), inserindo procedimentos geralmente utilizados em outras áreas (por exemplo, eólica), tais como o uso de um maior número de indicadores de qualidade (flags), que permitem obter uma melhor classificação dos dados, e o uso da técnica estatística box whiskers (boxplot). Além dos testes comumente utilizados (testes locais), ainda são propostos testes globais que avaliam a série temporal como um todo.

DESCRIÇÃO DOS LOCAIS

Para a aplicação do procedimento proposto, foram escolhidos dois locais no Brasil. O primeiro, Petrolina, é um município do estado de Pernambuco, situado no Nordeste brasileiro. Possui coordenadas geográficas (-9,06; -40,31), altitude de 387 m e clima semiárido quente (BSh). O segundo, Brasília, capital do país, situa-se no Distrito Federal, Centro-Oeste brasileiro. Possui coordenadas geográficas (-15,60; -47,71), altitude de 1023 m e clima tropical com estação seca (Aw). Ambos os locais apresentam dados coletados pela rede SONDA – Sistema de Organização Nacional de Dados Ambientais. Os dados são constituídos pela irradiância solar global, direta e difusa, possuindo intervalo de integração de 1 minuto. Para ambas as regiões, foi utilizado um ano de dados – 2011 para Brasília e 2014 para Petrolina. A escolha destes anos se deu após uma análise preliminar dos dados, que mostrou a ausência de componentes da radiação solar em outros anos nestes locais, fato que impossibilita a total utilização do procedimento de garantia de qualidade apresentado neste trabalho.

METODOLOGIA

Indicadores de qualidade (flags)

A utilização de indicadores – também chamados de flags – nos procedimentos de garantia de qualidade é de extrema importância para caracterizar o nível de confiança de determinada amostra de uma série temporal. A forma mais comumente utilizada é a classificação binária, onde determinado dado será classificado apenas como dado bom ou dado anômalo. Neste trabalho, optou-se por utilizar um número maior de indicadores de qualidade que o usual, a fim de se obter uma melhor caracterização dos dados. Cada teste do procedimento de garantia de qualidade carrega consigo uma matriz coluna de flags associada, na qual cada amostra da série temporal é classificada com um, e apenas um, dos números listados abaixo na Tabela 1. Se um dado for classificado como “bom” (flag 1), significa que ele foi aprovado em todos os testes do procedimento de garantia de qualidade anteriores ao teste em análise. Um dado é classificado como “suspeito” (flag 2) quando ele falha em determinado teste cujo teor não é essencialmente conclusivo quanto à classificação do dado, ou seja, quando tal teste não nos permite concluir se uma amostra é fruto de uma medição correta ou é um espúrio, também chamado de *outlier*. Com isso, ao final do processo, caso um dado suspeito não seja classificado como anômalo em nenhum teste posterior, fica a cargo do usuário a utilização desse dado em seu projeto. Se um dado for classificado como “anômalo” (flag 3) em determinado teste, ele não será avaliado em nenhum dos testes posteriores, sendo classificado como dado “previamente anômalo” (flag 4) em todos os testes subsequentes. A classificação “dado não testado” (flag 5) indica que determinada amostra não será avaliada em algum teste do procedimento. Por exemplo, neste trabalho

são classificados como não testados todos os dados de radiação solar que representam o período da noite, além do início e fim do dia solar. Por fim, o indicador “dado não disponível” (flag 6) é utilizado para identificar que o dado bruto de um determinado instante de tempo não foi registrado pelo sistema de aquisição de dados – SAD.

Indicador	Descrição
1	Dado bom
2	Dado suspeito
3	Dado anômalo
4	Dado previamente anômalo
5	Dado não testado
6	Dado não disponível

Tabela 1: Descrição dos indicadores de qualidade

Abordagem não paramétrica (Box Whiskers)

O diagrama de caixa, box whiskers ou boxplot é uma técnica estatística utilizada para a determinação de limites superior e inferior de aceitação para classificar um dado como potencialmente bom ou potencialmente anômalo. A abordagem é denominada “não paramétrica” por não depender da natureza da variável utilizada, fazendo com que este procedimento possa ser estabelecido em vários testes. Com isso, a técnica utiliza-se do conceito de quartis e percentis para determinar a concentração de dados em um histograma. Tendo-os em mãos, os limites superior e inferior serão determinados pelo intervalo abaixo:

$$[Q1 - k * IQR ; Q3 + k * IQR] \quad (1)$$

Os termos à esquerda e à direita do intervalo representam os limites inferior e superior, respectivamente. Q3 e Q1 são referentes ao terceiro e primeiro quartil da distribuição respectivamente. IQR é o intervalo interquartil, que pode ser representado por $IQR = (Q3 - Q1)$. O parâmetro k é chamado de multiplicador. Seu valor usualmente é utilizado como $k = 1,5$, assim como visto em Younes et al. (2005). Porém, um problema comumente enfrentado é o ajuste do procedimento para distribuições assimétricas, com elevada obliquidade. Nesse caso, o IQR se aproxima de zero, diminuindo os valores para os limites superior e inferior e rechaçando uma grande quantidade de dados. A fim de sanar este problema, Hubert e Vandervieren (2007) propuseram uma nova metodologia para o uso do diagrama em distribuições com elevada obliquidade, incluindo um termo exponencial nas equações descritas em (1). Neste termo, se faz presente uma função estatística chamada *Medcouple* (MC), que é utilizada para calcular a obliquidade de distribuições univariadas e é empregada tal como definem Brys et al. (2004). Com isso, o intervalo (1) será substituído pelo intervalo (2) ou (3) dados por:

$$[Q1 - k * e^{-4MC} * IQR ; Q3 + k * e^{3MC} * IQR], \text{ para } MC \geq 0 \text{ e } k = 1,5 \quad (2)$$

$$[Q1 - k * e^{-3MC} * IQR ; Q3 + k * e^{4MC} * IQR], \text{ para } MC < 0 \text{ e } k = 1,5 \quad (3)$$

Uma característica do método de Hubert e Vandervieren (2007) é o estabelecimento de limites muito extensos, distantes da média da distribuição. Porém, uma possível desvantagem do método é quando tais limites são extrapolados de maneira que se perca o sentido no que tange o comportamento

fisicamente esperado da variável em estudo. Por isso, é proposta uma união do procedimento estabelecido por Hubert e Vandervieren (2007) com o procedimento empregado por Moraes (2015), o qual se baseia em Eischeid et al. (1995), Baker (1994) e Bussab e Morettin (2006) para a obtenção do multiplicador k . O método consiste no cálculo da porcentagem de rechaço “P” para vários valores de k discretizados ao longo do domínio R_+ . Com isso, uma curva (dP) da diferença de primeira ordem de P é traçada e o ponto de “mínimo” é localizado. O valor do multiplicador “ k ” será aquele associado ao ponto de estabilização (ponto de mínimo) da curva dP .

Testes da garantia de qualidade

Os testes adotados no procedimento de garantia de qualidade devem ser executados sequencialmente, tal como mostra a Figura 1. Isto se deve ao fato de que os primeiros testes visam eliminar as anomalias mais bruscas, baseando-se em limites físicos de aceitação para as variáveis em questão. Caso isto não fosse feito, tais amostras anômalas exerceriam grande influência nos testes subsequentes, uma vez que modificariam as distribuições estatísticas envolvidas (o que conseqüentemente exerceria influência na estimativa dos limites superiores e inferiores de aceitação para as variáveis).



Figura 1: Testes do procedimento de garantia de qualidade.

Testes globais

Os testes globais são aplicados à série temporal como um todo visando avaliar se a mesma apresenta inconsistências, tais como, descontinuidade, dados ausentes, lacunas e falhas no registro de data e hora. De acordo com o resultado dos testes globais, a série pode ser descartada ou seguir adiante para a próxima etapa do processo (testes locais). Os testes globais utilizados são aplicados nas séries das radiações global, direta e difusa e estão descritos a seguir.

a) Descontinuidade

Por determinação da EPE – Empresa de Pesquisa Energética, uma série de dados de no mínimo 2 anos que contenha uma lacuna contínua de dados equivalente a um período de 15 dias não deve ser considerada por inteiro. Para não inutilizar a série de medição, a mesma pode ser subdividida em séries temporais independentes (uma anterior e outra posterior à lacuna).

b) Missing data

Segundo a EPE, uma série de medição só poderá ser considerada válida para utilização caso possua no máximo 10% de dados ausentes (flag 6). Uma verificação dessa condição é realizada para as séries de radiação solar global, direta e difusa.

c) Duplicidade de datas

Na série de dados fornecida pelo instrumento de medição, encontram-se informações sobre o instante de tempo em que ocorreu a medição de cada amostra. Caso existam, na mesma série, dois instantes de tempo exatamente iguais, desconsideram-se as amostras referentes a estes dois instantes de tempo.

d) *Lacunas*

A série temporal dispõe de amostras que distam entre si (duas amostras sucessivas) de um certo intervalo de tempo “ t ”, no qual cada instante de medição é fornecido pelo SAD. Caso duas amostras consecutivas possuam um intervalo de tempo maior que “ t ”, entre estas amostras está caracterizada uma lacuna temporal. Para sanar este problema, propõe-se preencher as lacunas com NaNs (*Not a Number*), no intuito de preservar a sequência temporal e evitar a associação errônea entre dois dados de duas classes diferentes naquele instante de tempo – por exemplo, observar a global e a componente direta correspondente.

Testes locais

Os testes locais são testes que visam a avaliação de amostras ou conjunto de amostras da série temporal, utilizando-se tanto de modelos estatísticos, que não dependem da natureza da variável (modelos não paramétricos), quanto modelos físicos baseados na dinâmica da radiação solar em suas componentes global, direta e difusa. Os testes utilizados neste trabalho são descritos a seguir juntamente com as referências correspondentes.

a) *Ângulo de elevação*

De acordo com Maxwell et al. (1993), dados de radiação solar cujo ângulo de elevação $\alpha \leq 10^\circ$ deverão ser descartados por estarem sujeitos a uma faixa de incerteza dos sensores de medição. Com isso, o teste será aplicado a todas as séries temporais e todos os dados compreendidos nesse intervalo são classificados como “dado não testado”, atribuindo o flag 5.

b) k_t, k_d, k_n

Nos testes de um procedimento de garantia de qualidade, é desejável que se eliminem primeiramente os *outliers* mais grosseiros, que fogem completamente dos demais dados. Assim, sendo I_g a irradiância global, I_{bn} a irradiância direta normal, I_{0eff} a irradiância extraterrestre e θ_z o ângulo zenital, definem-se $k_t = I_g / [I_{0eff} \cdot \cos(\theta_z)]$, $k_d = I_d / I_g$ e $k_n = I_{bn} / I_{0eff}$. Com isso, propõe-se neste trabalho que os limites de aceitação para valores de k_t , k_d e k_n sejam os primeiros filtros, dentre todos os filtros físicos adotados. Younes et al. (2005) adotam que $0 < k_t < 1$ e $0 < k_d < 1$. Raichijk (2012), por sua vez, estabelece um intervalo maior para os valores de k_d ($0 < k_d < 1,1$) devido à incerteza dos sensores utilizados. Dessa forma, considerando-se as incertezas dos sensores de medição, optou-se por escolher os limites de $0 < k_t < 1,1$; $0 < k_d < 1,1$ e $0 < k_n < 1,1$ atribuindo o flag 3 aos dados de irradiância global, direta e difusa que ultrapassarem tais limites.

c) *BSRN*

A BSRN propõe um conjunto de testes físicos para I_g , I_{bn} e I_d , estabelecendo limites superiores com base na irradiância extraterrestre. Assim, limites físicos são propostos para as irradiâncias global, direta normal e difusa e classificados como fisicamente possíveis (flag 3 em caso de falha) e extremamente raros (flag 2 em caso de falha), cujas equações seguem abaixo:

Limites fisicamente possíveis:

$$-4 \leq I_g \leq I_{0\text{eff}} * 1,5 * \cos(\theta_z)^{1,2} + 100 \quad (4)$$

$$-4 \leq I_{\text{bn}} \leq I_{0\text{eff}} \quad (5)$$

$$-4 \leq I_d \leq I_{0\text{eff}} * 0,95 * \cos(\theta_z)^{1,2} + 50 \quad (6)$$

Limites extremamente raros:

$$-2 \leq I_g \leq I_{0\text{eff}} * 1,2 * \cos(\theta_z)^{1,2} + 50 \quad (7)$$

$$-2 \leq I_d \leq I_{0\text{eff}} * 0,75 * \cos(\theta_z)^{1,2} + 30 \quad (8)$$

$$-2 \leq I_{\text{bn}} \leq I_{0\text{eff}} * 0,95 * \cos(\theta_z)^{0,2} + 10 \quad (9)$$

d) Modelo de céu claro

O modelo proposto no atlas solarimétrico europeu (ESRA) para representar a irradiância de céu claro em suas componentes global ($I_{g,cc}$), direta ($I_{b,cc}$) e difusa ($I_{d,cc}$), é utilizado para estabelecer limites máximos para os valores medidos de irradiância solar. Os valores máximos aceitos para cada componente de radiação não devem ultrapassar em até 10% a radiação de céu claro para a global e a direta. Já para a difusa, a radiação de céu claro não pode ultrapassar em 20% a radiação medida. A tolerância maior para a difusa foi proposta devido ao fato que dias com índices de claridade elevados ou médios apresentam baixos valores de I_d . Essas medições tendem a apresentar maiores erros em razão de se situarem na região de maior incerteza na faixa de operação do sensor utilizado. Para a simulação do modelo ESRA, foi considerado o valor de Turbidez de Linke igual a 2,5, conforme trabalho de Ruiz-Arias et al. (2010). Os dados que falharem em seus respectivos testes receberão flag 3.

e) Tracker off

De acordo com Long e Shi (2006), se $I_d > 50 \text{ W/m}^2$, $\theta_z < 75^\circ$ e as condições abaixo ocorrerem ao mesmo tempo, o tracker não está seguindo corretamente o sol. Caso isso ocorra, o flag 3 será atribuído para I_{bn} e I_d .

$$\frac{I_d + I_{\text{bn}} * \cos(\theta_z)}{I_{g,cc}} > 0,85 \quad (10)$$

$$\frac{I_d}{I_d + I_{\text{bn}} * \cos(\theta_z)} > 0,85 \quad (11)$$

Um erro no tracker é facilmente observável pelo comportamento mostrado na Figura 1, visto que os dados marcados em vermelho possuem um elevado valor de k_t associados a baixos valores de k_n . Tal comportamento não ocorre em condições naturais da radiação solar.

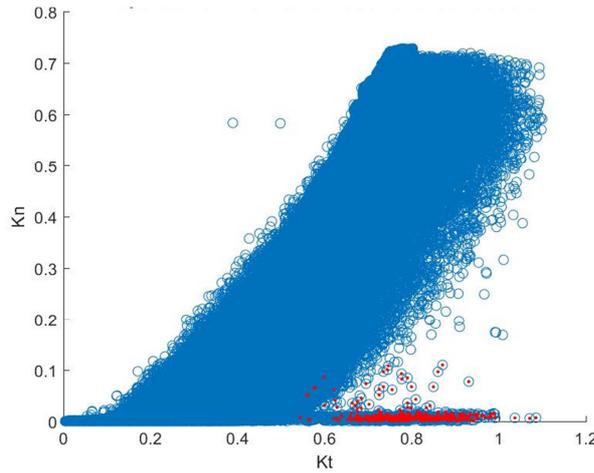


Figura 2: Dispersão $k_t - k_n$, aplicada à Petrolina, com amostras que falharam no teste “Tracker off” em vermelho.

f) *Consistência de parâmetros*

A BSRN propõe dois testes para analisar a consistência entre parâmetros derivados da radiação solar em suas componentes global, direta e difusa. O primeiro, impõe limites mínimos e máximos de variação para a razão entre a irradiância global medida I_g e a irradiância global calculada $I_d + I_{bn} * \cos(\theta_z)$. O segundo impõe limites mínimos e máximos de variação para o k_d . Um terceiro teste estabelece uma comparação entre valores de k_t e k_d e foi proposto por Raichijk (2012). Como não é possível identificar qual (quais) componente (s) apresenta (m) anomalia, as componentes envolvidas receberão flag 2 em caso de falha do teste.

$$\left| \frac{I_d + I_{bn} * \cos(\theta_z) - I_g}{I_g} \right| \leq 0,08 \text{ para } \theta_z < 75^\circ \text{ e } I_d + I_{bn} * \cos(\theta_z) > 50 \text{ W/m}^2 \quad (12)$$

$$\left| \frac{I_d + I_{bn} * \cos(\theta_z) - I_g}{I_g} \right| \leq 0,15 \text{ para } 75^\circ < \theta_z < 93^\circ \text{ e } I_d + I_{bn} * \cos(\theta_z) > 50 \text{ W/m}^2 \quad (13)$$

$$k_d < 1,05 \text{ para } \theta_z < 75^\circ \text{ e } I_g > 50 \text{ W/m}^2 \quad (14)$$

$$k_d < 1,1 \text{ para } 75^\circ < \theta_z < 93^\circ \text{ e } I_g > 50 \text{ W/m}^2 \quad (15)$$

$$k_d \geq 0,9 \text{ para } k_t < 0,2 \quad (16)$$

$$k_d \leq 0,8 \text{ para } k_t > 0,6 \quad (17)$$

g) *Teste de envelope*

Younes et al. (2005) sugerem um teste de envelope que consiste de uma dispersão entre valores de k_t e k_d , onde o eixo que contém os valores de k_t é dividido em 10 bins (intervalos). São calculados a média e o desvio padrão dos pontos contidos em cada bin. Após isso, para cada intervalo, são

estabelecidos os limites $[\bar{k}_d - 2\sigma_{k_d}, \bar{k}_d + 2\sigma_{k_d}]$, que serão representados por dois pontos (um para cada limite) em cada bin, totalizando no final 20 pontos ao longo da distribuição – levando em conta que temos 10 bins –, sendo 10 relativos a limites superiores e 10 relativos a limites inferiores. Com isso, Younes et al. (2005) propõem que sejam ajustados polinômios de segundo grau aos pontos superiores e inferiores, fazendo com que todos os pontos que estejam acima da curva superior ou abaixo da curva inferior sejam considerados anômalos. Uma vez que a medição da radiação difusa se mostra bastante problemática em várias regiões por motivos citados anteriormente, decidiu-se utilizar k_n ao invés de k_d , adotando os mesmos procedimentos utilizados por Younes et al. (2005), agora para uma dispersão entre k_t e k_n . Os dados anômalos recebem o flag 3.

h) Consistência temporal

O teste de consistência temporal visa estabelecer um limite aceitável para a variação na magnitude da radiação solar entre amostras sucessivas ao longo do tempo. A WMO – *World Meteorological Organization*, propõe testes de controle de qualidade para duas classes de dados. A primeira é relativa à dados brutos, no qual o procedimento de controle de qualidade é realizado em estações meteorológicas automáticas e os testes visam eliminar erros grosseiros. A segunda diz respeito a dados processados, na qual os testes aplicados possuem um maior nível de refinamento, sendo utilizados após eliminação de erros grosseiros. Segundo a WMO, o teste de consistência temporal possui um limite máximo de 800 W/m² para dados brutos e 1000 W/m² para dados processados. Nesse trabalho, adota-se o limite de 1000 W/m², uma vez que os *outliers* mais bruscos já foram previamente eliminados pelos testes anteriores. Com isso, amostras consecutivas que diferirem entre si de mais de 1000 W/m² serão consideradas anômalas, com ambas recebendo flag 3.

i) Persistência

Mal funcionamento do sensor, quedas de energia ou sombreamentos constantes podem fazer com que sejam retornados dados idênticos em sequência durante determinado tempo. O teste de persistência visa identificar um comprimento máximo aceitável para repetições de valores idênticos ao longo do tempo. Para isso, é proposta a utilização da metodologia de Box Whiskers. Como a radiação solar possui comportamentos distintos ao longo do dia, propõe-se a divisão dos dados em intervalos definidos pela massa de ar (MA), como foi feito em Raichijk (2012). Assim, os dados contidos nos intervalos $1 < MA < 1,25$, $1,25 < MA < 2,5$ e $2,5 < MA < 5,76$ serão separados em diferentes distribuições para posterior análise. Além disso, como o comportamento da radiação difere bastante em céu claro e céu nublado, para cada intervalo serão criados dois subconjuntos. O primeiro é relativo aos dados cujo $k_t < 0,7$ naquele instante. O segundo, relativo aos dados cujo $k_t > 0,7$. Após isso, é empregada a metodologia de Box Whiskers para cada conjunto de distribuições obtidos. Uma vez que a distribuição possui uma elevada obliquidade, se utilizará da metodologia proposta anteriormente neste trabalho. Com isso, os dados que ultrapassarem os limites superiores estabelecidos para cada distribuição serão considerados anômalos e receberão flag 3.

RESULTADOS

Testes Globais

Os conjuntos de dados das duas localidades não falharam em nenhum dos testes globais, permitindo concluir que ambas as séries temporais podem ser devidamente filtradas pelos testes locais para posterior utilização.

Testes Locais

Abaixo se encontram os resultados dos testes aplicados à Petrolina e Brasília, no qual serão exibidas a quantidade percentual atribuída a cada flag, em cada teste do procedimento, assim como a quantidade

percentual de rechaço final, ambas com relação à quantidade total de dados de cada estação (525.600 para ambas). A disponibilidade dos dados para cada componente será indicada pelo flag 6. Com isso, separa-se a análise em três “blocos” de filtros, no qual cada bloco possui características similares que serão discutidas abaixo.

a) “Ângulo de elevação”, “ k_i, k_d, k_n ”, “BSRN” e “Modelo de céu claro”:

Ângulo de elevação							kt, kd, kn					
Petrolina			Brasília				Petrolina			Brasília		
Flag	Ig	Ibn	Id	Ig	Ib	Id	Ig	Ibn	Id	Ig	Ib	Id
1	44,1	44,1	44,1	42,5	42,5	42,5	44,1	41,7	44,0	42,5	39,2	42,4
2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,4	0,1	0,0	3,3	0,1
4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
5	55,9	55,9	55,9	56,1	56,1	56,1	55,9	55,9	55,9	56,1	56,1	56,1
6	0,0	0,0	0,0	1,4	1,4	1,4	0,0	0,0	0,0	1,4	1,4	1,4

BSRN						Modelo de céu claro						
Petrolina			Brasília			Petrolina			Brasília			
Flag	Ig	Ibn	Id	Ig	Ib	Id	Ig	Ibn	Id	Ig	Ib	Id
1	44,1	41,7	44,0	42,5	39,2	42,4	43,5	41,7	43,0	41,7	39,2	37,4
2	0,0	0,0	1,2	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0
3	0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,0	1,0	0,9	0,0	5,0
4	0,0	2,4	0,1	0,0	3,3	0,1	0,0	2,4	0,2	0,0	3,3	0,1
5	55,9	55,9	55,9	56,1	56,1	56,1	55,9	55,9	55,9	56,1	56,1	56,1
6	0,0	0,0	0,0	1,4	1,4	1,4	0,0	0,0	0,0	1,4	1,4	1,4

Tabela 2: Resultados percentuais dos testes “Ângulo de elevação”, “ k_i, k_d, k_n ”, “BSRN” e “Modelo de céu claro” para Petrolina e Brasília.

Para o primeiro bloco, os testes aplicados carregam consigo o intuito de eliminar as anomalias mais grosseiras, que fogem do comportamento fisicamente esperado das variáveis em questão. Nota-se que tais testes exerceram papel fundamental no filtro de I_{bn} e I_d para ambas as localidades. No entanto, para I_g , menos de 1% dos dados falharam nos testes para ambas as localidades. Isto se deve ao fato da maioria dos dados de I_g que falham em tais testes encontrarem-se em instantes de tempos para os quais $\alpha < 10^\circ$, o que os caracterizam como dados não testados (flag 5). A dificuldade de tais testes identificarem anomalias em I_g indica a importância dos testes subsequentes (menos utilizados que os primeiros pelo público em geral) no resultado final.

b) “Tracker off”, “Consistência de parâmetros” e “Teste de envelope”:

Tracker off							Consistência de parâmetros						Teste de envelope					
Petrolina			Brasília				Petrolina			Brasília			Petrolina			Brasília		
Flag	Ig	Ibn	Id	Ig	Ib	Id	Ig	Ibn	Id	Ig	Ib	Id	Ig	Ibn	Id	Ig	Ib	Id
1	-	40,4	41,4	-	39,1	37,3	41,1	39,9	39,0	38,5	36,1	34,2	40,0	36,9	-	37,5	34,9	-
2	-	0,0	0,0	-	0,0	0,0	2,4	0,5	2,4	3,2	3,0	3,2	0,0	0,0	-	0,0	0,0	-
3	-	1,3	1,6	-	0,1	0,1	-	-	-	-	-	-	3,6	3,6	-	4,2	4,2	-
4	-	2,4	1,1	-	3,3	5,1	0,6	3,7	2,7	0,9	3,4	5,2	0,6	3,7	-	0,9	3,4	-
5	-	55,9	55,9	-	56,1	56,1	55,9	55,9	55,9	56,1	56,1	56,1	55,9	55,9	-	56,1	56,1	-
6	-	0,0	0,0	-	1,4	1,4	0,0	0,0	0,0	1,4	1,4	1,4	0,0	0,0	-	1,4	1,4	-

Tabela 3: Resultados percentuais dos testes “Tracker off”, “Consistência de parâmetros” e “Teste de envelope” para Petrolina e Brasília.

Os testes utilizados neste bloco são de natureza comparativa, no qual se utilizam dois ou mais componentes da radiação solar. É importante notar que tais testes exerceram papel fundamental na filtragem de I_g , identificando o conjunto de anômalos finais em quase sua totalidade, além de participarem efetivamente na identificação de anômalos de I_{bn} . É importante notar a quantidade percentual de dados classificados como suspeitos no teste de consistência de parâmetros. Uma

dispersão entre a irradiância global medida (I_g) e a irradiância global calculada ($I_{bn} \cdot \cos(\theta_z) + I_d$) mostra claramente o peso dos dados suspeitos na análise final, tal como mostra a Figura 3.

c) “Consistência temporal” e “Persistência”:

Flag	Consistência temporal						Persistência					
	Petrolina			Brasília			Petrolina			Brasília		
	Ig	Ibn	Id	Ig	Ib	Id	Ig	Ibn	Id	Ig	Ib	Id
1	40,0	36,8	-	37,5	34,9	-	40,0	36,9	39,0	37,5	34,9	34,2
2	0,0	0,0	-	0,0	0,0	-	0,0	0,0	0,0	0,0	0,0	0,0
3	0,0	0,0	-	0,0	0,0	-	0,0	0,0	0,0	0,0	0,0	0,0
4	4,1	7,3	-	5,0	7,6	-	4,1	7,3	2,7	5,0	7,6	5,2
5	55,9	55,9	-	56,1	56,1	-	55,9	55,9	55,9	56,1	56,1	56,1
6	0,0	0,0	-	1,4	1,4	-	0,0	0,0	0,0	1,4	1,4	1,4

Tabela 4: Resultados percentuais dos testes “Consistência temporal” e “Persistência” para Petrolina e Brasília.

Para o terceiro e último bloco, apresentam-se testes que possuem a função de eliminar as anomalias cujo comportamento não foi identificado em testes anteriores, tornando o procedimento de garantia de qualidade mais criterioso. Percebe-se que, para ambas localidades, nenhum dado das componentes envolvidas falhou nos testes. Isto indica que os testes anteriores obtiveram um bom desempenho na qualificação dos dados.

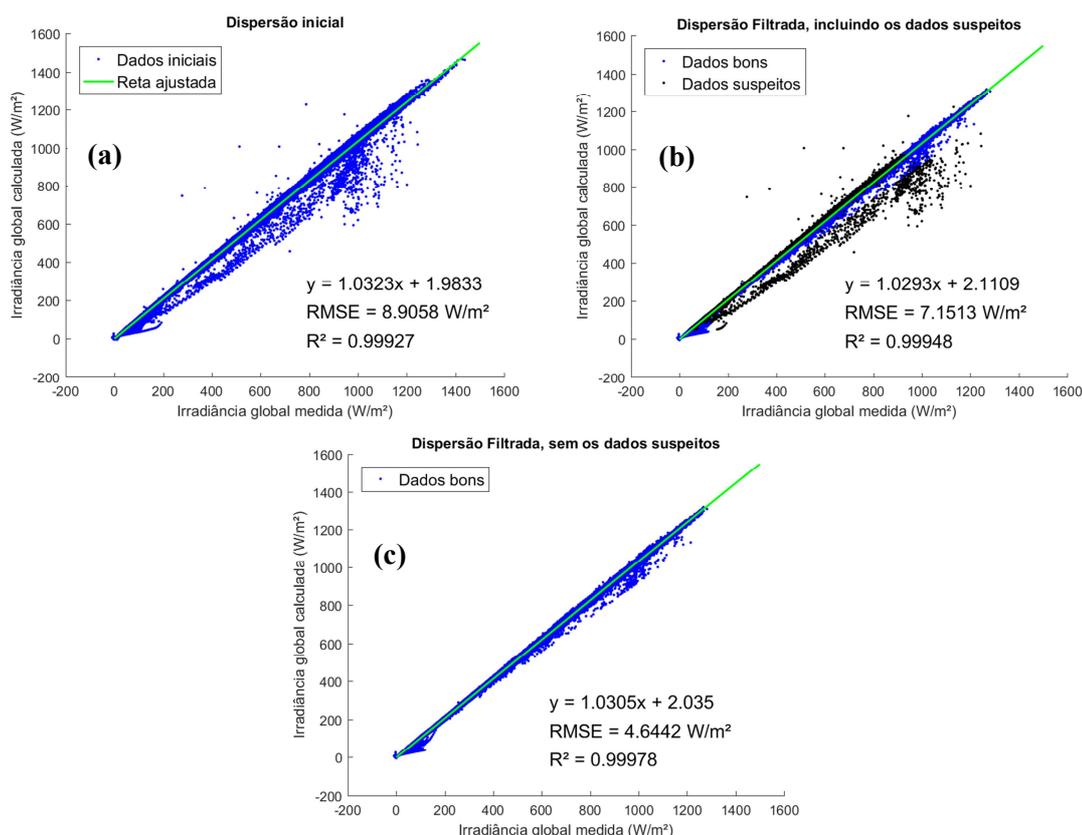


Figura 3: Dispersão da I_g medida vs I_g calculada para Petrolina - PE, considerando todos os dados (a) apenas dados bons e suspeitos (b) e apenas os dados bons (c).

Término do procedimento

Uma vez classificados os dados, cabe ao usuário decidir quanto a sua utilização. A tabela 5 mostra que, para ambas localidades, menos de 10% dos dados foram considerados anômalos para todas as

componentes. Tal percentual pode ser ultrapassado caso o usuário opte pela não utilização dos dados suspeitos (considerando-os como anômalos). Para o auxílio em sua decisão, gráficos como os da Figura 3 devem ser analisados pelo usuário. Observa-se que, apesar de não exercerem grande influência no coeficiente de determinação (R^2), os dados suspeitos influenciam significativamente na raiz do erro quadrático médio (RMSE) em uma dispersão entre a I_g medida e a I_g calculada (como mostrada anteriormente), promovendo uma redução de RMSE de aproximadamente 7 W/m² para cerca de 4 W/m² ao serem desconsiderados. Este resultado evidencia que, em certos casos, a utilização de dados suspeitos pode conferir aos dados incertezas indesejadas a depender do emprego dos mesmos, como, por exemplo, previsões de irradiância a curto prazo, sensíveis à comportamentos de alta frequência (característicos de séries com muitas anomalias). Tal fato evidencia a importância da qualificação dos dados.

Flags	Petrolina			Brasília		
	Global	Direta	Difusa	Global	Direta	Difusa
1	37,6	36,4	37,8	34,3	31,9	34,1
2	2,4	0,5	3,6	3,2	3,0	3,2
3	4,1	7,2	2,7	5,0	7,6	5,2
4	-	-	-	-	-	-
5	55,9	55,9	55,9	56,0	56,0	56,0
6	0	0	0	1,4	1,4	1,4

Tabela 5: Resultados percentuais finais para Petrolina e Brasília.

CONCLUSÃO

O procedimento utilizado de garantia de qualidade para dados de radiação solar se mostrou positivo para o que foi proposto. O fato de ser um procedimento automatizado contribui substancialmente no intuito de facilitar o processo de utilização dos dados de radiação solar. A união de vários procedimentos já difundidos na literatura juntamente com a proposição de pequenos ajustes e novos métodos entregou resultados aceitáveis, rechaçando as anomalias mais grosseiras e até os dados cujo comportamento anômalo não é notável a vista, os quais dificilmente seriam percebidos em uma análise subjetiva. No teste de persistência, a técnica de Box Whiskers, apesar de não ter acusado nenhuma anomalia, serve para aumentar o grau de confiabilidade do procedimento. Além disso, sua natureza não paramétrica permite a utilização do teste em bases de dados que não disponham das três componentes da radiação solar, fazendo com que tal teste possa ser utilizado em outros procedimentos de qualificação de dados a serem propostos, substituindo (quando necessário) testes de natureza comparativa – como, por exemplo, o teste “Tracker off”. Nos resultados finais, a série de Petrolina mostrou resultados levemente melhores que a de Brasília. Ambas as séries tiveram menos de 10% dos dados rechaçados para as componentes global, direta e difusa. Porém, notou-se que a inclusão dos dados suspeitos (considerando-os como anômalos) aumentaria tal percentual de rechaço. Após a classificação dos dados, uma sugestão ao usuário é a aplicação de um procedimento de preenchimento de dados ao final do processo (caso as anomalias sejam descartadas), visando a diminuição (quando possível) das lacunas da série final.

REFERÊNCIAS

- Baker, N. L. (1994). Quality Control of Meteorological Observations at Fleet Numerical Meteorology and Oceanography Center, NAVAL RESEARCH LAB MONTEREY CA.
- Brys, G., Hubert, M. e Struyf, A. (2004). A robust measure of skewness. *Journal of Computational and Graphical Statistics* 13(4): 996-1017.
- Bussab, W. O. e Morettin, P. A. (2006). *Estatística Básica*, Saraiva.
- Eischeid, J. K., Baker, C. B., Karl, T. R. e Diaz, H. F. (1995). The Quality Control of Long-Term Climatological Data Using Objective Data Analysis. *Journal of Applied Meteorology* 34(12): 2787-2795.

- Gilchrist, B. e Cressman, G. P. (1954). An experiment in objective analysis. *Tellus* 6(4): 309-318.
- Hubert, M. e Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational statistics & data analysis* 52(12): 5186-5201.
- Journée, M. e Bertrand, C. (2011). Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy* 85(1): 72-86.
- Long, C. N. e Shi, Y. (2006). The QCRad value added product: Surface radiation measurement quality control testing, including climatology configurable limits. Atmospheric Radiation Measurement Program Technical Report.
- Maxwell, E., Wilcox, S. e Rymes, M. (1993). Users manual for SERI QC software, assessing the quality of solar radiation data. Solar Energy Research Institute, Golden, CO.
- Moraes, C. F. W. C. B. (2015). Procedimento Objetivo para a Garantia da Qualidade de Dados Observacionais de Vento em Superfície no Litoral do Rio Grande do Norte. Mestrado, UFPE.
- Raichijk, C. (2012). Control de calidad de mediciones de radiación solar. Presentado en XXXV Reunión de Trabajo de la Asociación Argentina de Energías Renovables y Ambiente (ASADES), Rosario, Argentina.
- Ruiz-Arias, J., Alsamamra, H., Tovar-Pescador, J. e Pozo-Vázquez, D. (2010). Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management* 51(5): 881-893.
- Sönmez, İ. (2013). Quality control tests for western Turkey Mesonet. *Meteorological Applications* 20(3): 330-337.
- Younes, S., Claywell, R. e Muneer, T. (2005). Quality control of solar radiation data: present status and proposed new approaches. *Energy* 30(9): 1533-1549.

ABSTRACT: In order to enable the use of reliable data, it is proposed an objective procedure for the quality assurance of solar radiation data. This work establishes the union of already diffused procedures in the literature (physical and statistical tests) – along with some adjustments – and the use of unpublished procedures in the quality assurance of solar radiation data (e.g. box whiskers). To obtain the results, the procedure was applied on data from the SONDA network, in stations located at Petrolina – PE and Brasília – DF, both at Brazil. All the time series had less than 10% of rejected solar radiation data.

Keywords: quality assurance; solar radiation; global tests, box whiskers.