

Data-QS-Ferramenta numérica para qualificação de dados coletados em estações automáticas de superfície**Data-QS-Numerical tool for qualification of data collected in automatic surface stations**

DOI:10.34117/bjdv5n7-143

Recebimento dos originais: 14/06/2019

Aceitação para publicação: 16/07/2019

Thaiane Gambarra Soares

Bacharel em Engenharia de Petróleo e Recursos Renováveis
Instituto do Mar – Universidade Federal de São Paulo – *campus* Baixada Santista
Rua Carvalho de Mendonça, 144. Encruzilhada, Santos/SP. CEP: 11070-100

Christine Camargo Mendes

Bacharelanda em Engenharia de Petróleo e Recursos Renováveis
Instituto do Mar – Universidade Federal de São Paulo – *campus* Baixada Santista
Rua Carvalho de Mendonça, 144. Encruzilhada, Santos/SP. CEP: 11070-100

Glaucia Clososki

Bacharelanda em Engenharia Ambiental
Instituto do Mar – Universidade Federal de São Paulo – *campus* Baixada Santista
Rua Carvalho de Mendonça, 144. Encruzilhada, Santos/SP. CEP: 11070-100

Gabriel Tonello Souza

Bacharelando em Engenharia Ambiental
Instituto do Mar – Universidade Federal de São Paulo – *campus* Baixada Santista
Rua Carvalho de Mendonça, 144. Encruzilhada, Santos/SP. CEP: 11070-100

André Luiz Vizine Pereira

Doutor em Ciências da Computação e Matemática Computacional pela Universidade de São Paulo.

Instituto do Mar – Universidade Federal de São Paulo – *campus* Baixada Santista
Rua Carvalho de Mendonça, 144. Encruzilhada, Santos/SP. CEP: 11070-100

William Remo Pedroso Conti

Doutor em Física pelo Instituto de Física da Universidade de São Paulo
Instituto do Mar – Universidade Federal de São Paulo – *campus* Baixada Santista
Rua Carvalho de Mendonça, 144. Encruzilhada, Santos/SP. CEP: 11070-100

Fernando Ramos Martins

Doutor em Geofísica Espacial pelo Instituto Nacional de Pesquisas Espaciais
Instituto do Mar – Universidade Federal de São Paulo – *campus* Baixada Santista
Rua Carvalho de Mendonça, 144. Encruzilhada, Santos/SP. CEP: 11070-100

RESUMO

O Data-QS (Data Qualification System) é uma ferramenta computacional desenvolvida em linguagem de programação Python para qualificação e análise preliminar de dados coletados por sensores meteorológicos e radiométricos operando em estações automáticas em superfície. O controle de qualidade de dados é um requisito fundamental para estudos de viabilidade técnica e econômica do aproveitamento do recurso renovável de energia. O controle de qualidade de dados segue critérios recomendados pela World Meteorological Organization. O pacote Data-QS, além de informar a presença de dados suspeitos na base de dados observados, produz análises gráficas das séries temporais. O programa está em fase final de desenvolvimento e aprimoramento da interface gráfica e os algoritmos de qualificação de dados foram testados com uso da base de dados disponibilizados pela rede SONDA.

Palavras-chave — Qualificação de dados meteorológicos, Sensoriamento remoto da atmosfera, Avaliação de recursos energéticos, Energia Solar, Energia Eólica.

ABSTRACT

The Data-QS (Data Qualification System) is a computational tool developed in Python for quality check of meteorological and radiometric data acquired by in automated weather stations. The quality-check procedure is a fundamental requirement for technical and economical evaluation of renewable power plants. The data-quality criteria used in Data-QS are based on the World Meteorological Organization (WMO) recommendations for automated weather stations. Besides data quality evaluation, Data-QS also generates reports and graphical analysis of the observational data series. Data-QS is in the final phase of adjustments and improvement of interface was developed and tested by using the observational data acquired in SONDA measurement sites as open access.

Key words — Quality control of meteorological data, Remote sensing of atmosphere, Energy resource assessment, Solar energy, Wind energy.

1 INTRODUÇÃO

A segurança energética é um tema estratégico no momento atual em razão da sua relação intrínseca não apenas com a infraestrutura necessária para dar suporte ao desenvolvimento socioeconômico do país, mas também com a qualidade de vida da sociedade. No entanto, a preocupação ambiental ganhou relevância nas últimas décadas em razão do fato que o consumo de energia é apontado com um dos fatores de maior contribuição ao aumento da concentração de gases de efeito estufa na atmosfera e, por consequência, com as mudanças climáticas (Tolmasquim *et al.*, 2007). Neste contexto, diversos países têm realizado um esforço para desenvolver tecnologias e estabelecer metas para adoção de fontes renováveis de energia.

A matriz elétrica brasileira está baseada principalmente em fontes renováveis de energia – aproximadamente 75% da eletricidade gerada (Tolmasquim *et al.*, 2007). No entanto,

a geração hidroelétrica concentra grande parte desta contribuição sendo responsável por 64% da oferta interna de energia elétrica (Agência Nacional de Energia Elétrica, 2018). A participação da energia eólica alcançou cerca de 8% da matriz elétrica nacional, enquanto que a participação da energia solar se restringe a 1,2 % (Agência Nacional de Energia Elétrica, 2018). A concentração na geração hidroelétrica resulta em aumento da vulnerabilidade do sistema elétrico nacional em razão de prolongados períodos de estiagem.

Desde 2002, o Governo Brasileiro vem adotando políticas de incentivo à diversificação da matriz elétrica por meio de aproveitamento de outras fontes renováveis, como a energia eólica e a solar. O PROINFA foi o principal programa de incentivo na última década e resultou no crescimento do aproveitamento do recurso da energia dos ventos principalmente no Nordeste brasileiro, onde os períodos de estiagem afetam de forma significativa os reservatórios das usinas hidrelétricas (Pereira *et al.*, 2017). Recentemente, a geração solar fotovoltaica também está crescendo em um ritmo acelerado em razão da redução de custos da tecnologia e em decorrência da regulamentação adotada pelo Governo para a geração distribuída e a realização de leilões específicos para compra de eletricidade gerada a partir da conversão da energia solar (EPE, 2018).

A avaliação de disponibilidade de recursos de fontes renováveis de energia requer a coleta e análise de dados observados em campo em razão da relação intrínseca entre esses recursos e as condições meteorológicas e climatologia do local pretendido para a instalação da planta de geração (Tolmasquim *et al.*, 2007). A sazonalidade e a intermitência dessas fontes precisam ser conhecidas a fim de minimizar as incertezas e riscos nos investimentos realizados.

Os recursos de energia solar e eólica possuem características bastantes peculiares e uma intermitência associada aos ciclos sazonais decorrentes do movimento orbital da Terra em torno do Sol, além de fatores locais como relevo e uso do solo. Para avaliar a disponibilidade e a variabilidade temporal desses recursos, é necessário instalar e operar estações para coleta de dados meteorológicos, incluindo sensores para medida de irradiação solar, velocidade e direção do vento que atendam padrões técnicos mínimos para reduzir as incertezas associadas com a aquisição de dados de campo (Tolmasquim *et al.*, 2007). Além da qualidade e a instalação correta dos equipamentos utilizados nas estações, é fundamental manter um controle de qualidade dos dados observados com o intuito de garantir a representatividade dos dados em relação às condições meteorológicas locais e possibilitar a análise de comparação com diferentes bases de dados para a mesma localidade.

O controle de qualidade dos dados é adotado internacionalmente e está baseado em critérios recomendados pela Organização Meteorológica Mundial (OMM) e por organizações voltadas para o setor energético (WMO, 2004). A rede internacional BSRN (Baseline Surface Radiation Network) de coleta de dados radiométricos é administrada pela WMO e possui um padrão bastante consolidado para o controle de qualidade de dados observados (WRMC, 2018). O Brasil possui algumas estações de superfície que integram a BSRN e, portanto, devem atender os padrões de qualidade estabelecidos para a rede (Pereira *et al.*, 2017).

A apresentação de propostas de plantas de geração elétrica para os leilões de energia realizados no Brasil devem conter uma base de dados observados das variáveis meteorológicas básicas por um período mínimo de um ano atendendo um padrão de qualidade semelhante ao estabelecido para a BSRN (EPE, 2018). Os critérios de qualidade estabelecidos incluem análise da variabilidade temporal de variáveis meteorológicas e comparação com valores fisicamente possíveis e extremamente raros. Dada a complexidade dos critérios estabelecidos e a extensão da série de dados, o uso de ferramentas numéricas é fundamental para realização do controle de qualidade dos dados observados de forma automatizada e com confiabilidade na identificação dos padrões de qualidade não atendidos.

A ferramenta computacional apresentada neste artigo foi desenvolvida para controle de qualidade de dados meteorológicos, incluindo dados de irradiação solar, velocidade e direção dos ventos. O Data-QS foi desenvolvido com uso da linguagem Python e auxílio de ferramentas de software livre e sem custos. O pacote está em fase final de desenvolvimento e será disponibilizado livremente para uso público.

O desenvolvimento dos algoritmos para o controle de qualidade dos dados, em conjunto com este artigo, desempenham um papel importante na conscientização e capacitação de recursos humanos para a busca de qualidade e confiabilidade de dados coletados em campo a fim de garantir avaliações consistentes sobre a disponibilidade de recursos renováveis e viabilidade técnica-econômica da implantação do aproveitamento do recurso natural.

2 MATERIAIS E MÉTODOS

O procedimento de controle de qualidade de dados tem como objetivo principal identificar registros que possam apresentar incertezas elevadas e/ou sejam decorrentes de problemas técnicos durante a operação nas estações de coleta de dados operando em superfície. O método é idêntico para qualquer localização geográfica da estação de coleta de dados. A metodologia aqui descrita tem como ponto de partida o procedimento de controle de qualidade estabelecido pela *World Meteorological Organization* (WMO) para identificação de dados espúrios em estações da *Baseline Surface Radiation Network* (BSRN) que é uma rede de referência para coleta de dados ambientais e meteorológicos (Pereira *et al.*, 2017).

Inicialmente, a base de dados deve ser verificada quanto à integridade da sequência temporal dos dados registrados no arquivo. Por exemplo, a base de dados deve seguir uma sequência cronológica válida, sem repetições ou ausências de dados. Quando falhas na sequência temporal de aquisição ocorrem, o código computacional deve informar a falha na aquisição ou duplicidade de registro em períodos de tempo específicos no relatório de qualificação de dados. Caberá ao proprietário da base de dados em avaliação ou usuário do *software*, verificar a origem do problema e optar por realizar as devidas correções manualmente com a inclusão dos valores observados nos horários faltantes ou fazer uso de bandeira (*flag*) para indicação de falha. O *software* também checa a ocorrência de registros específicos de um dos sensores em operação na estação em avaliação num dado horário ou período de coleta de dados.

Após o teste de integridade, o processo de qualificação dos dados é realizado em 3 etapas sequenciais nas quais os dados são sinalizados como suspeitos de acordo com as premissas apresentadas na Tabela 1. As etapas são sequenciais, ou seja, somente após a avaliação e aprovação em uma etapa é que o registro coletado será submetido para a seguinte. Se o dado for considerado suspeito em qualquer uma das etapas, o processo não terá continuidade para este que será sinalizado como suspeito.

Tabela 1. Etapas do algoritmo utilizado para avaliação de qualidade de dados observados em estações de superfície.

Etapa	Variáveis Anemométricas	Variáveis Meteorológicas	Variáveis Radiométricas
1	Dado suspeito quando fisicamente impossível		
2	Dado suspeito quando extremamente raro		

3	Dado suspeito quando a variabilidade temporal da variável em análise não condiz com o esperado fisicamente.	Dado suspeito se não houver consistência em comparação com valores de outras variáveis coletadas na mesma estação.
---	---	--

Cada registro das variáveis observadas nas estações de coleta de dados é avaliado e o resultado da qualificação é apresentado na forma de um código numérico de 4 dígitos (*flag* ou indicador de qualidade) que caracteriza todo o procedimento de qualificação. Este indicador deve ser lido da direita para esquerda de modo que o primeiro dígito à direita se refere à primeira etapa do procedimento de qualificação na etapa 1, o segundo - da direita para esquerda - se refere à etapa 2 e assim sucessivamente. Cada dígito pode receber um dos 4 valores numéricos distintos que identificam a qualidade do registro observado, conforme descrito na Tabela 2.

As Tabelas 3 e 4 listam os critérios adotados para cada uma das variáveis monitoradas em uma estação padrão de coleta de dados exigida pela ANEEL (EPE, 2018) para credenciamento de projetos de geração de eletricidade para os leilões de energia.

Tabela 2. Valores atribuídos a cada uma das quatro posições do indicador de qualidade dos registros observados nas estações de coleta de dados.

Resultado do teste de qualidade	Descrição da significado físico
0	Nenhum procedimento foi executado
2	Dado suspeito
5	Procedimento não pode ser executado
9	Dado aprovado

Para compreender os critérios listados na Tabela 4, é necessário assumir que AZS representa o ângulo zenital solar; μ_0 representa o cosseno do ângulo zenital solar; S_0 é a constante solar na distância média entre Terra e Sol; UA é a correção de excentricidade da órbita da Terra; $S_a = S_0/UA^2$ representa o valor da constante solar ajustada para a distância entre Sol-Terra. A variável GLO corresponde a irradiação solar global incidente em uma superfície horizontal medida por um piranômetro sem sombreador enquanto que DIF é a irradiação solar difusa horizontal medida por um piranômetro com sensor sombreado para evitar a incidência direta do feixe de radiação solar. A variável DNI corresponde à componente direta da

irradiação solar com incidência normal na superfície e *Sum* corresponde a soma das componentes difusa (*DIF*) e direta incidente perpendicularmente à uma superfície horizontal ($DNI \cdot \mu_0$). A constante de Stephan-Boltzmann está representada por σ . A temperatura do ar (em Kelvin) está representada por T_a ; e LW_{dn} é a irradiação de onda longa descendente medida por um pirgeômetro.

O código computacional para todas as etapas de qualificação de dados foi desenvolvido em Python com uso das bibliotecas Numpy e Pandas. A interface gráfica para interação com o usuário (GUI – *Graphic User Interface*) foi desenvolvida com uso da biblioteca nativa Tkinter. Os relatórios disponibilizados após a execução dos testes de qualidade são apresentados em forma de tabelas e gráficos gerados através da biblioteca Matplotlib.

Tabela 3. Critérios de qualidade adotados para avaliação para as variáveis meteorológicas.

Etapa 1	Etapa 2	Etapa 3
Temperatura (T)		
$T_{\min.}$ e $T_{\max.}$: com base em valores típicos para climatologia local	$\Delta T < 5^\circ \text{C}$ em um período de 1h	$\Delta T > 0,5^\circ \text{C}$ em um período de 12h consecutivas
Umidade Relativa (UR)		
$UR_{\min.}$: 20 % $UR_{\max.}$: 100 %	$\Delta UR \leq 1\%$ em um minuto	$\Delta UR \geq 1\%$ em período de 1h consecutiva
Pressão Atmosférica (p)		
$p_{\min.}$ e $p_{\max.}$: com base em valores típicos para climatologia local	$\Delta p \leq 0,5 \text{ mb}$ em observações consecutivas	$\Delta p \geq 0,1 \text{ mb}$ em intervalo de 1h
Precipitação (R)		
$R_{\min.}$: 0 mm $R_{\max.}$: típico para climatologia local	$\Delta R < 25 \text{ mm}$ em 1h consecutiva	$\Delta R < 100 \text{ mm}$ em um período de 24h consecutivas

Tabela 4. Critérios de qualidade adotados para avaliação dos registros das variáveis radiométricas.

Etapa 1	Etapa 2	Etapa 3
Global Horizontal (GLO)		
mín.: - 4 W/m ² máx.: $1,5.Sa. \mu_0^{1,2} + 100$ (W/m ²)	mín.: -2 W/m ² máx.: $1,2.Sa. \mu_0^{1,2} + 50$ (W/m ²)	se $Sum > 50 \text{ W/m}^2$ e $AZS < 75^\circ$: $GLO/Sum = \pm 10\%$ se $Sum > 50 \text{ W/m}^2$ e $AZS > 75^\circ$: $GLO/Sum = \pm 15\%$ se $Sum < 50 \text{ W/m}^2$: não realizado
Direta Normal (DNI)		
mín.: -4 W/m ² máx.: Sa (W/m ²)	mín.: -2 W/m ² máx.: $0,95.Sa. \mu_0^{0,2} + 10$ (W/m ²)	$DNI. \mu_0 - 50 \leq (GLO - DIF) \leq$ $\leq DNI. \mu_0 +$ 50
Difusa (DIF)		
mín.: -4 W/m ² máx.: $0,95.Sa. \mu_0^{1,2} + 50$ (W/m ²)	mín.: -2 W/m ² máx.: $0,75.Sa. \mu_0^{1,2} + 30$ (W/m ²)	se $GLO > 50 \text{ W/m}^2$ e $AZS < 75^\circ$: $DIF/GLO < 1,05$ se $GLO > 50 \text{ W/m}^2$ e $75^\circ < AZS < 93^\circ$: $DIF/GLO < 1,1$ se $GLO < 50$: não realizado
Onda Longa (LW_{dn})		
mín.: 40 W/m ² máx.: 700 W/m ²	mín.: 60 W/m ² máx.: 500 W/m ²	$0,4.T_a^4.LW_{dn} < T_a^4 + 25$ (W/m ²)

3 RESULTADOS

A Figura 1 apresenta a tela inicial do *software* Data-QS. Para execução dos testes, o usuário deverá fornecer o caminho para um arquivo de dados, em formato “.csv”, contendo uma planilha com os dados observados na estação de superfície. Caso os dados armazenados na planilha estejam organizados de maneira idêntica ao padrão estabelecido para a rede SONDA (<http://sonda.ccst.inpe.br>), o Data-QS identifica automaticamente as variáveis meteorológicas em cada coluna após a leitura do arquivo. Caso contrário, o usuário deverá indicar cada uma das variáveis a serem testadas pelo *software* e sua respectiva coluna. Após o preenchimento das informações pelo usuário, o Data-QS realiza uma etapa de pré-processamento necessária para a organização dos dados e identificação de falhas na série temporal disponibilizada para análise no arquivo de dados.

The screenshot shows the 'Data-QS Software' interface with the 'Análise' (Analysis) tab selected. The interface is divided into several sections for data input and processing:

- Arquivo modelo da rede SONDA?**: Radio buttons for 'Sim' (Yes) and 'Não' (No).
- Tipo de qualificação**: Radio buttons for 'Solarimétrico', 'Anemométrico', and 'Meteorológico'.
- Estação meteorológica mais próxima:** A text input field.
- Estações Meteorológicas**: A list of meteorological stations with an 'Upload' button.
- Versão dos dados:**: Radio buttons for 'Anterior a 3.3' and 'A partir de 3.3'.
- Insira a coluna correspondente:**: A list of variables with checkboxes: Timestamp, Ano, Dia, Minuto, Radiação Global Horizontal, Radiação Direta, Radiação Difusa, Radiação de Onda Longa, Radiação PAR, Iluminância, Temperatura do Ar na Superfície, Umidade do ar, Pressão Atmosférica, Precipitação da Chuva, Velocidade do Vento, and Direção do Vento.
- Insira os valores ou escolha uma estação próxima:**: Input fields for Latitude, Longitude, Altitude, Temp max, Temp Min, and Precip Max.
- Buttons**: 'Run' and 'Exportar' (Export) buttons.

Figura 1. Interface inicial com o usuário do pacote computacional Data-QS para entrada de informações necessárias para avaliação da qualidade de dados observados em estações de coleta de dados por sensoriamento remoto.

Após o pré-processamento dos dados, os testes baseados nos critérios de qualidade são realizados. Ao término, o Data-QS grava dois arquivos em formato “.csv”. O primeiro armazena uma planilha com os dados reorganizados em padrão similar aos dados da rede SONDA. Os registros de qualidade, para cada um dos dados presentes, estarão armazenados no segundo arquivo de igual tamanho e formatação. A ferramenta computacional também fornece um relatório descrevendo um resumo da qualidade da base de dados como um todo – porcentagem de dados suspeitos em cada variável considerada e análise gráfica com base nos parâmetros indicados no preenchimento na interface da ferramenta computacional no momento da entrada de dados.

A Figura 2 apresenta um exemplo ilustrativo de relatório gráfico produzido pelo software Data-QS. O relatório completo é gerado com as informações sobre a fração de dados de cada uma das variáveis meteorológicas e/ou radiométricas identificadas como com qualidade suspeita em cada etapa do procedimento de controle de qualidade da base de dados.

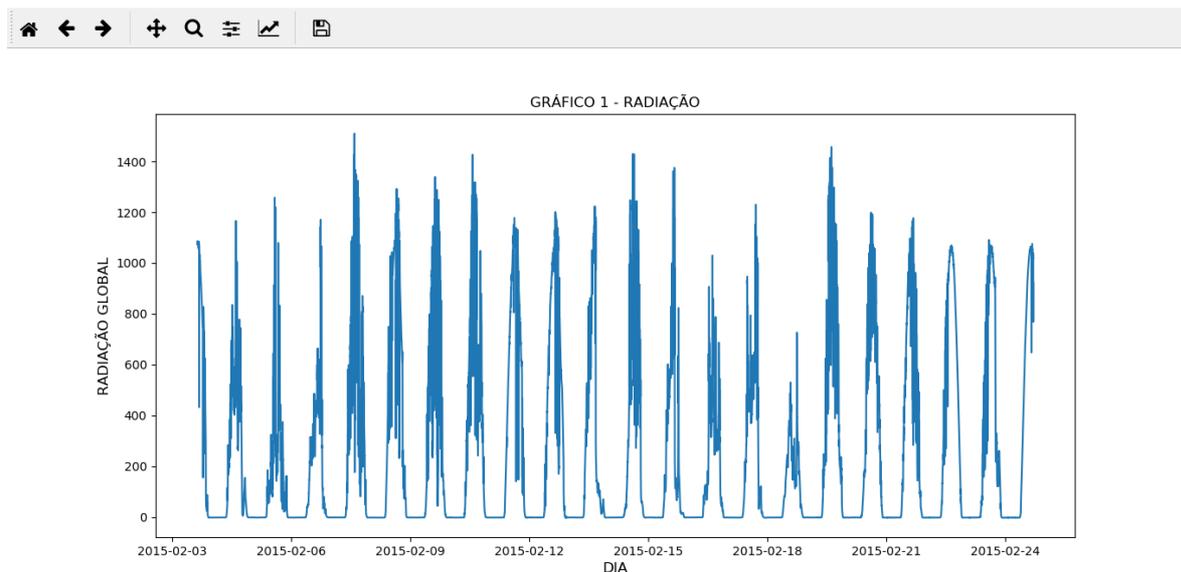


Figura 2. Relatório gráfico fornecido pelo software Data-QS para série temporal de irradiação global incidente na superfície (em W/m^2) em função da data de aquisição dos dados.

5. CONCLUSÕES

O Laboratório Interdisciplinar de Computação Aplicada (LAICA) tem por missão o desenvolvimento de ferramentas computacionais composto por docentes e discentes do campus Baixada Santista da Universidade Federal de São Paulo. O LAICA visa atender demandas apresentadas pela sociedade nas diversas áreas das atividades socioeconômicas. Essa atuação tem propiciado aos estudantes um espaço de aprimoramento e capacitação em técnicas de programação e de interação com a sociedade para promover uma formação de bacharéis e engenheiros com formação humanista e socialmente referenciada.

O software Data-QS é uma ferramenta computacional para avaliação de controle de qualidade de dados meteorológicos e radiométricos observados em estações automáticas de aquisição de dados e destinados ao monitoramento do recurso energético solar e eólico para o empreendedores do setor de energia. Os critérios adotados estão cientificamente estabelecidos com base em instituições internacionais consolidadas na área de sensoriamento remoto da atmosfera. A ferramenta computacional estará disponível para uso livre e sem custos para o público em geral.

A ferramenta computacional Data-QS é de uso livre e gratuito para toda a comunidade e tem como objetivo promover o uso de informações cientificamente validadas para avaliação de recursos de fontes renováveis de energia. O Data-QS é produto de trabalho de um grupo de estudantes envolvidos nas atividades de extensão do Laboratório Interdisciplinar em Computação Aplicada da UNIFESP *campus* Baixada Santista.

AGRADECIMENTOS

À UNIFESP pela infraestrutura e auxílio de bolsa de extensão para estudantes dos cursos de Engenharia de Petróleo e Recursos Renováveis e Engenharia Ambiental. Agradecimentos ao LabREN/INPE pela cessão de dados observados na rede SONDA para o desenvolvimento e testes do Data-QS.

REFERÊNCIAS

Agência Nacional de Energia Elétrica. Banco de dados de Geração (BIG): Capacidade de Geração. <online>: <http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>. 2018.

Atlas Brasileiro de Energia Solar. LABREN / CCST / INPE (Instituto Nacional de Pesquisas Espaciais). <online>: http://labren.ccst.inpe.br/atlas_2017.html , 2017.

Empresa de Pesquisa Energética (EPE). Instruções para cadastramento de projetos para participar em leilões de energia. <http://www.epe.gov.br/pt/leiloes-de-energia/instrucoes-para-cadastramento>, 2018.

Pereira *et al.* Atlas Brasileiro de Energia Solar. 2^a. Ed. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2017.

Mauricio Tolmasquim, Amilcar Guerreiro e Ricardo Gorini. Matriz Energética Brasileira: Uma Prospectiva, 2007.

World Radiation Monitoring Center (WRMC). Baseline Surface Radiation Network – Project and Database. <online>: <https://bsrn.awi.de>, 2018.

World Meteorological Organization (WMO). Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. CBS/OPAG-IO/ET AWS-3/Doc. 4(1), 2004.