

UMA ANÁLISE ESTATÍSTICA COM EMPREGO DE TÉCNICAS DE REGRESSÃO LINEAR MÚLTIPLA PARA A SUAUIZACÃO DA VARIABILIDADE DA GERAÇÃO SOLAR FOTOVOLTAICA EM LARGA-ESCALA CONECTADA AO SISTEMA INTERLIGADO NACIONAL (SIN)

Gabriel Dantas de Oliveira Rolim – biel.dantasrolim@gmail.com
Leonardo José de Petribú Brennand – leojdpb@hotmail.com
Valentin Paschoal Perruci – vpperruci@gmail.com
Luiz Homero Câmara Medeiros – lhcm@uol.com.br
Olga de Castro Vilela– ocv.olga@gmail.com
Alexandre Carlos Araújo da Costa – alex.araujodacosta@gmail.com

Filiação: Universidade Federal de Pernambuco, Departamento de Energia Nuclear (DEN-UFPE)

Resumo. *A natureza intermitente da geração de energia eólica e solar, cujas participações na matriz elétrica nacional têm aumentado significativamente ao longo dos últimos anos (BEN, 2016), impulsiona a criação de métodos e ferramentas mais sofisticados para a operação da rede elétrica nacional. Simultaneamente, é de fundamental importância considerar a disposição espacial e a potência instalada de cada uma das plantas, com vistas a aumentar a estabilidade do despacho desses recursos intermitentes e com alto grau de incerteza no Sistema Interligado Nacional (SIN). Neste contexto, este trabalho apresenta um método simples e eficaz, que faz uso do modelo de regressão linear múltipla, para o estabelecimento da potência nominal de plantas fotovoltaicas de larga escala (potência da ordem de dezenas de MW), de tal forma a utilizar a complementariedade do recurso solar dos diversos locais estudados em prol do aumento da suavização (smoothness) e da previsibilidade da potência elétrica gerada. Para este estudo foram levados em consideração 16 locais de interesse contidos no Nordeste brasileiro, região com o maior potencial solar disponível (ANEEL, 2002). Os dados que descrevem o comportamento do recurso solar de cada um desses locais em escala horária foram obtidos a partir da saída de modelos físicos disponibilizados pelo NSRDB (National Solar Radiation Database).*

Palavras-chave: *Energia Solar Fotovoltaica; Regressão Linear Múltipla; Sistema Interligado Nacional*

1. INTRODUÇÃO

Existe uma preocupação básica a todas as concessionárias e autoridades governamentais do setor elétrico em diversos países com relação ao crescente grau de inserção das energias eólica e solar na matriz elétrica, que antes havia acomodado níveis iniciais dessas fontes primárias de energia intermitente sem impactos significativos na operação do sistema interligado de potência, em termos de parâmetros de confiabilidade, segurança e disponibilidade.

Uma das formas mais indicadas para minimizar a intermitência da potência produzida por fontes renováveis como solar e eólica é fazendo uso da possível complementariedade do recurso disponível em diversos locais (Hoicka e Rowlands, 2011). Ao observar estudos anteriores dedicados à minimização da intermitência do recurso solar (Mills e Wisser, 2010), foco deste trabalho, é perceptível que a minimização da intermitência é abordada apenas através da localização geográfica das centrais, não levando em consideração a minimização da intermitência através do uso de centrais com diferentes potências nominais. Obviamente, a escolha da localização das centrais é de fundamental importância, visto que a complementariedade entre as centrais só será efetiva se tais centrais estiverem situadas em locais com diferentes características climáticas. No entanto, ao utilizarmos centrais com mesma potência nominal, ou desconsiderarmos a potência nominal das centrais a serem instaladas, não estaremos usufruindo eficientemente da complementariedade do recurso solar disponível.

O objetivo deste trabalho se traduz na apresentação de uma metodologia, que faz uso de regressão linear múltipla, para ponderar as potências nominais de plantas fotovoltaicas de larga escala situadas em 16 locais da região nordeste no Brasil, de tal forma a compor uma configuração favorável ao despacho da potência elétrica no Sistema Interligado Nacional, isso é, uma configuração em que a complementariedade do recurso solar dos diversos locais estudados é utilizada mais eficientemente com vistas a aumentar a suavidade e previsibilidade da potência elétrica despachada.

O artigo será dividido em quatro seções ao todo, sendo esta a primeira. A Seção 2 apresentará os dados utilizados para realizar este estudo e a metodologia empregada. A Seção 3 discutirá sobre os resultados obtidos, os quais se apresentam ainda em fase de desenvolvimento, sendo, portanto, de caráter preliminar. Por fim, a Seção 4 apresentará as conclusões, considerações finais e perspectivas de emprego da metodologia proposta.

2. METODOLOGIA E MODELO

A metodologia apresentada neste trabalho tem como objetivo estabelecer critérios de distribuição da potência nominal de centrais solares de grande porte através da aplicação de técnicas baseadas em regressão linear múltipla (conforme Geladi e Kowalski, 1986 e Hong, 2010), as quais são ajustadas para descrever o comportamento de uma central resultante, com menos intermitência, em função de 16 ou mais plantas solares de larga escala conectadas ao Sistema Interligado Nacional (SIN).

2.1. Descrição dos dados

Os dados solarimétricos utilizados neste trabalho foram obtidos da base de dados NSRDB (National Solar Radiation Database), pertencente ao laboratório americano NREL (National Renewable Energy Laboratory). Os dados têm resolução temporal horária e advêm de um modelo físico (PSM) que estima as irradiâncias global, direta e difusa através de dados de satélite (GOES) em conjunto com outros modelos (PATMOS-x, REST2, etc) responsáveis por modelar as condições atmosféricas, possuindo, por fim, uma resolução espacial de 4 x 4 km. Foram obtidos dois anos de dados (2013 e 2014) para 16 locais, situados no Nordeste brasileiro. Além disso, a irradiância de céu claro (I_{CS}) foi estimada, para todos os locais de estudo, de acordo com Rigollier *et al.* (2000).

2.2. Locais de interesse

Foram escolhidos 16 locais com variadas características climáticas, todos situados na região Nordeste do Brasil, local de estudo deste trabalho. O critério de escolha se deu principalmente pela distância entre os locais, com o intuito de observar o efeito da suavização espacial pela penetração de potência fotovoltaica de cada local na rede (Mills e Wiser, 2010). A distância em linha reta entre os locais varia de cerca de 60 km a cerca de 1200 km. Informações relevantes dos locais estudados são mostradas na Tab. 1.

Tabela 1 – Informações relevantes dos locais estudados.

LOCAL	LATITUDE	LONGITUDE	ELEVAÇÃO (ASL)
Bom Jesus – PI	-9,07	-44,35	430 m
Bom Jesus da Lapa – BA	-13,24	-43,41	273 m
Brumado – BA	-14,19	-41,66	415 m
Caicó – RN	-6,45	-37,09	153 m
Floresta – PE	-8,59	-38,56	311 m
Garanhuns – PE	-8,89	-36,49	841 m
Irecê – BA	-11,29	-41,85	718 m
Itabaianinha – SE	-11,27	-37,78	220 m
Mossoró – RN	-5,18	-37,34	20 m
Natal – RN	-5,79	-35,21	38 m
Ouricuri – PE	-7,87	-40,07	434 m
Patos – PB	-7,02	-37,27	221 m
Petrolina – PE	-9,39	-40,50	380 m
Piatã – BA	-13,15	-41,76	1258 m
Serra Talhada – PE	-7,98	-38,29	443 m
Teresina – PI	-5,08	-42,80	87 m

2.3. Suavização da variabilidade fotovoltaica

Para mitigar a variabilidade fotovoltaica no sistema de potência, é preciso entender e analisar as causas de tal fenômeno. A variabilidade na geração está intimamente ligada à variação da radiação solar incidente no plano coletor. A radiação solar, por sua vez, é regida por duas vertentes no tocante à sua intermitência. A primeira diz respeito à sua variação diária e sazonal, representada pela geometria Sol – Terra, facilmente prevista deterministicamente. A segunda leva em conta as atenuações atmosféricas e o sombreamento por nuvens, os quais são dificilmente previstos devido às suas características estocásticas. A última carrega consigo uma variabilidade solar não previsível, referida em algumas literaturas pelo termo “taxa de rampa” (Harper, 2010). Tal fenômeno mensura variações abruptas na geração fotovoltaica, muito danosas ao sistema de potência, que necessita do rápido acionamento da sua reserva girante ou geração térmica para compensar tal imprevisto.

Diante disso, podemos dividir a problemática da suavização fotovoltaica em dois enfoques distintos: o primeiro aborda a suavização do ponto de vista da complementariedade dos efeitos determinísticos, o segundo aborda a suavização do ponto de vista da complementariedade dos efeitos estocásticos.

A suavização da variabilidade da geração solar fotovoltaica em larga escala no cotidiano operacional do sistema de potência a partir da perspectiva de complementariedade de características determinísticas depende intrinsecamente da localização geográfica das centrais, principalmente suas longitudes, pois essa influencia diretamente na fase do sinal elétrico gerado. Portanto, para que possamos atingir uma grande homogeneidade no sinal elétrico gerado a partir de plantas fotovoltaicas é necessário distribuir as centrais ao longo de uma grande faixa de longitudes.

Considerando que o objetivo deste trabalho se concentra na suavização da variabilidade da geração solar fotovoltaica de centrais localizadas na Região Nordeste do Brasil, a faixa de longitudes disponível para a implantação de centrais se torna bastante restrita.

Em vista disso, a metodologia exposta no presente trabalho objetiva a minimização da variabilidade fotovoltaica a partir do ponto de vista da complementariedade dos efeitos estocásticos, de modo a tentar aproximar ao máximo a geração de energia elétrica de um conjunto de centrais fotovoltaicas, situadas nos locais de interesse, da geração de energia elétrica que seria obtida a partir de um modelo de céu claro. Dessa forma, a variabilidade da geração solar fotovoltaica seria mais bem comportada, tendo em vista que ela se aproximaria da variabilidade de um modelo determinístico bastante suave.

Uma vez que os locais de interesse possuem coordenadas geográficas muito próximas, o comportamento do modelo de céu claro em todos esses locais é muito semelhante, e esse fato pode ser verificado através do menor valor do coeficiente de correlação entre as séries do modelo de céu claro de quaisquer dois locais de interesse. Por essa razão, é possível utilizar a média das séries temporais dos modelos de céu claro de cada um dos locais de interesse como uma série temporal de modelo de céu claro representativo de toda a região que envolve esses locais.

Partindo do pressuposto, ainda que rudimentar, de que a energia elétrica gerada por uma planta fotovoltaica depende linearmente da radiação solar incidente sobre a planta, a energia elétrica obtida a partir da simples escolha de locais para implantação de centrais fotovoltaicas com potências nominais semelhantes não necessariamente resultaria em uma melhor aproximação da geração fotovoltaica obtida a partir de um modelo de céu claro, uma vez que a média da radiação solar de todos os locais de interesse não obrigatoriamente se aproximará da radiação solar descrita através do modelo de céu claro, ou seja, não é possível garantir uma boa eficiência no combate às intermitências oriundas do recurso solar.

Nesse âmbito, a metodologia descrita neste trabalho visa determinar qual a contribuição de cada uma das centrais na potência nominal total a ser instalada, de tal forma que a potência elétrica gerada por todas essas centrais se aproxime da potência elétrica que seria gerada pelo modelo de céu claro representativo da região de interesse. Logo, a potência elétrica gerada por todas as centrais pode ser dada pela Eq. 1.

$$P_{total} = \sum_{i=1}^N P_i \quad (1)$$

Sendo N o número total de centrais a serem instaladas, P_{total} a série temporal da potência total a ser instalada e P_i a série temporal da potência gerada pela central i . Uma vez que a intenção dessa metodologia é aproximar a potência total gerada pelas centrais daquela que seria gerada por um modelo de céu claro, temos na Eq. 2:

$$P_{CS} = \sum_{i=1}^N P_i + \varepsilon \quad (2)$$

Sendo P_{CS} a potência elétrica gerada pelo modelo de céu claro representativo de toda a região que envolve os locais estudados e ε o erro entre P_{CS} e a soma das potências das centrais.

Visto que o objetivo deste trabalho é apenas estimar a potência nominal de várias plantas a serem instaladas, a utilização do pressuposto de que a potência elétrica gerada em uma planta fotovoltaica varia linearmente com a radiação solar incidente é bastante razoável. Logo, temos na Eq. 3:

$$k_{CS} \cdot I_{CS} = \sum_{i=1}^N k_i \cdot I_i + \varepsilon \quad (3)$$

Sendo os coeficientes lineares k aqueles que relacionam a irradiância incidente, I , e a potência elétrica gerada. É importante perceber que o k_i depende linearmente da potência nominal do parque i , isso significa dizer que, um parque de 10 MW possui um k dez vezes maior que o de um parque de 1MW. Dessa forma o coeficiente k_{CS} é numericamente igual a soma de todos os coeficientes k_i , pois, para efeito de efeito comparativo, a potência nominal relacionada à irradiância do modelo de céu claro representativo da região de interesse deve ser numericamente igual a soma das

potências nominais de todos os parques. Com vistas a identificar a contribuição de cada uma das centrais, independentemente da potência nominal total a ser instalada, a Eq. 3 foi dividida pela soma dos k_i , que é numericamente igual à k_{CS} , resultando na Eq. 4.

$$I_{CS} = \sum_{i=1}^N \left(\frac{k_i}{k_{CS}} \right) \cdot I_i + \varepsilon \quad (4)$$

Com vistas a facilitar a compreensão da metodologia exposta, a fração (k_i/k_{CS}) foi denominada de α_i , como pode ser visto na Eq. 5.

$$I_{CS} = \sum_{i=1}^N \alpha_i \cdot I_i + \varepsilon \quad (5)$$

Dessa forma, temos que a contribuição da central j para a composição da potência nominal fotovoltaica total a ser instalada (c_j) será dada pela Eq. 6. Ou seja, a potência nominal a ser instalada em cada localidade irá depender apenas do c_j e da potência nominal fotovoltaica total a ser instalada, como pode ser visto na Eq. 7.

$$c_j = \frac{\alpha_j}{\sum_{i=1}^N \alpha_i} \quad (6)$$

$$Pot_{Nominal_j} = c_j \cdot Pot_{Nominal_{total}} \quad (7)$$

Onde $Pot_{Nominal_j}$ é a potência nominal do parque j e $Pot_{Nominal_{total}}$ é a potência nominal fotovoltaica total a ser instalada.

Para que a soma das potências de todas as plantas instaladas se aproxime o máximo possível da potência elétrica que seria gerada por uma irradiação de céu claro, iremos calcular o valor dos coeficientes α de tal forma a minimizar o valor do erro médio quadrático apresentado na Eq. 8. Essa forma de cálculo dos coeficientes é conhecida como regressão linear múltipla (Geladi e Kowalski, 1986).

$$MSE = \frac{1}{T} \sum_{t=1}^T (\varepsilon_t)^2 \quad (8)$$

Sendo T a quantidade total de passos de tempo das séries temporais utilizadas.

É importante salientar que, devido à pequena diferença de fase entre essas séries, dadas suas localizações geográficas, e também ao valor sempre positivo das séries temporais de radiação solar incidente, nenhum dos coeficientes estimados a partir da minimização do erro médio quadrático deverá possuir valor negativo.

Devido ao caráter modular das plantas fotovoltaicas de larga escala, isto é, essas plantas podem ser discretizadas como a união de várias plantas menores de potência nominal constante - aqui nomeadas de plantas mínimas - é também possível calcular a contribuição de cada um dos locais utilizando diretamente as séries temporais da potência que seria gerada pela planta mínima em cada um desses sítios, como mostrado na Eq. 9. A grande vantagem dessa abordagem seria não mais partir do pressuposto que a potência elétrica gerada em uma planta fotovoltaica varia linearmente com a radiação solar incidente, propiciando, dessa forma, uma maior acurácia e confiabilidade na diminuição da intermitência.

$$P_{CS_{P.mínima}} = \sum_{i=1}^N \alpha_i P_{P.mínima_i} + \varepsilon \quad (9)$$

Sendo $P_{CS_{P.mínima}}$ a potência elétrica gerada pela planta mínima quando a irradiância incidente no plano coletor for dada pelo modelo de céu claro e $P_{P.mínima_i}$ a potência elétrica gerada pela planta mínima quando a irradiância do local i incidir sobre o plano coletor.

Dado que o foco deste trabalho está na implantação de centrais fotovoltaicas de larga escala (ordem de grandeza de dezenas de MW), sempre que a potência nominal estimada de um dado local for inferior à potência nominal mínima de instalação (valor mínimo de potência nominal abaixo do qual considera-se desprezível a contribuição de uma determinada central frente à potência nominal total), esse local não mais será levado em consideração, e as contribuições dos demais locais serão recalculadas. Esse procedimento é realizado até que todas as centrais possuam potências nominais superiores à potência nominal mínima de instalação. É válido enfatizar que esse procedimento também pode ser realizado apenas utilizando as contribuições relativas de cada um dos locais e comparando-as com uma contribuição relativa mínima. No presente estudo foi utilizado como uma contribuição relativa mínima o valor de 1,25%.

2.4. Estratégia de validação cruzada

No presente trabalho, foi empregada uma estratégia de validação cruzada (Thornton e Running, 1999) com vistas à avaliação do desempenho do modelo em um período diferente daquele em que os parâmetros foram estimados. O período utilizado para a calibração dos parâmetros foi o ano de 2013 e o período utilizado para avaliar o desempenho do modelo foi o ano de 2014. Na calibração (ano de 2013) os parâmetros α_i foram estimados através da minimização do erro médio quadrático, conforme a Eq. 7. Para validação os coeficientes α_i calculados durante a calibração foram aplicados às séries no ano de 2014 para avaliação do desempenho do modelo com base na raiz do erro médio quadrático (RMSE).

3. RESULTADOS PRELIMINARES

O menor coeficiente de correlação entre as séries temporais de modelo de céu claro de dois locais de interesse distintos foi de 0,9802, o que representa um valor bastante alto. Constatando, dessa forma, que a média dessas séries constitui uma série de céu claro representativa de toda a região que envolve os locais de interesse.

A Fig. 1 mostra a estimativa, realizada utilizando o período de calibração (2013), da contribuição percentual de cada um dos locais de interesse na composição da potência fotovoltaica total. Como foi estabelecido um limite inferior de contribuição de 1,25%, os locais de Serra Talhada e Ouricuri foram desconsiderados para a implantação de centrais. Contudo, é válido salientar que isso se deve ao conjunto de locais selecionados para esse trabalho e, dessa forma, caso outro conjunto de locais de interesse seja utilizado, talvez Serra Talhada e Ouricuri possuam uma maior representatividade.

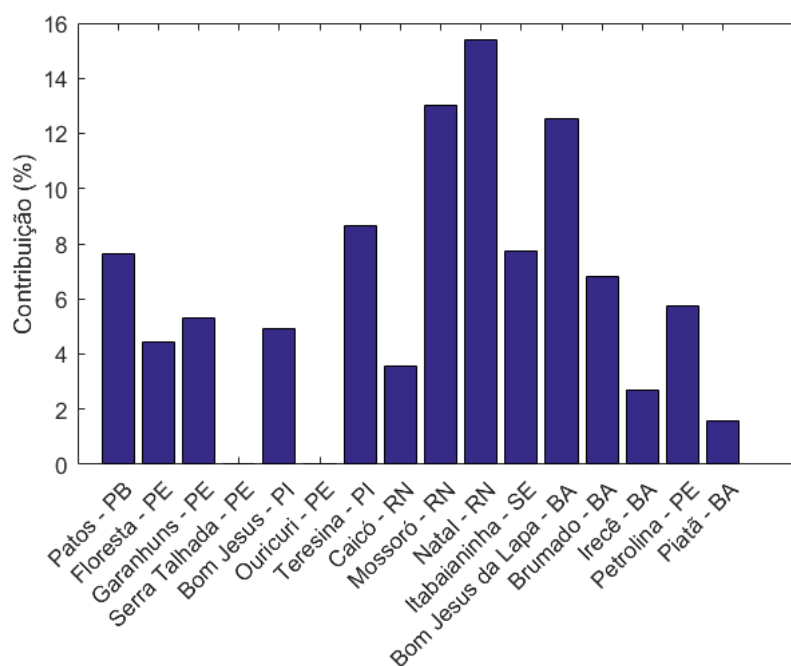


Figura 1- Contribuição de cada um dos locais de interesse na composição da potência fotovoltaica total.

A Fig. 2 mostra a raiz do erro médio quadrático, calculado utilizando o período de validação, entre a série do modelo de céu claro representativo de toda a região que engloba os locais de interesse e as séries estimadas por três modelos distintos, sendo eles: a regressão linear múltipla (MLR) utilizando os coeficientes estimados durante o período de calibração; a média das séries de irradiância de todos os locais de interesse (P-cte), que pode ser entendida como uma composição de centrais com mesma potência nominal; e a média das séries de irradiância dos locais com representatividade estimada na calibração superior à representatividade mínima (P-sel) (i.e. desconsiderando os locais de Serra Talhada e Ouricuri), que pode ser entendida como uma composição, utilizando a mesma potência nominal para todas as centrais, dos locais selecionados.

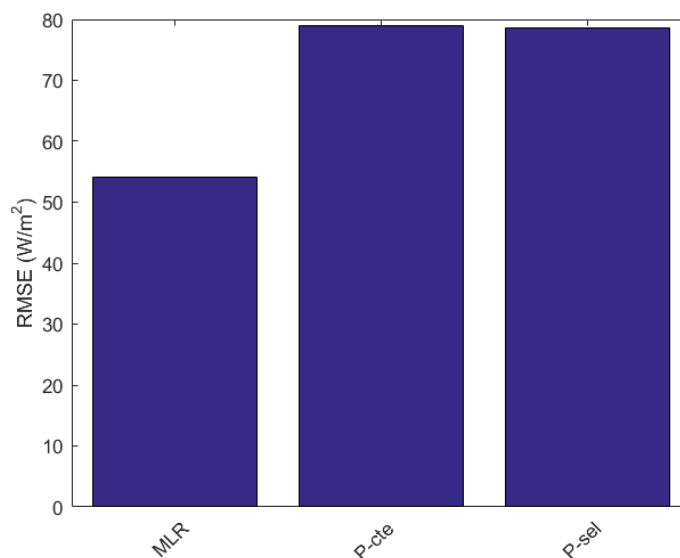


Figura 2- Raiz do erro médio quadrático durante o período de validação

Levando em consideração o pressuposto, ainda que rudimentar, de que a potência elétrica gerada pelas plantas fotovoltaicas pode ser considerada linearmente dependente da irradiância incidente, a diferença entre a raiz do erro médio quadrático do modelo MLR e os demais modelos para séries de irradiância demonstra o grande benefício em se utilizar plantas fotovoltaicas de locais distintos com potências nominais distintas para se obter uma maior suavização na variabilidade da potência elétrica total gerada.

4. CONCLUSÃO E PERSPECTIVAS

Os resultados apresentados neste trabalho mostram o grande potencial de melhoria em se utilizar plantas fotovoltaicas com potências nominais distintas para se obter uma maior suavidade na potência fotovoltaica total produzida, mesmo que as contribuições de cada uma das centrais sejam estimadas fazendo uso de modelos estatísticos de grande simplicidade.

Vale destacar que os dados de irradiância utilizados no atual trabalho, mesmo que bastante úteis para uma primeira abordagem e demonstração da metodologia proposta, são oriundos de modelagens físicas com resolução temporal horária. Contudo, para análises mais precisas, se faz necessária a utilização de dados observacionais com resolução temporal intra-horária, a fim de que seja possível a percepção de características microescalares dos locais de interesse, i.e. características oriundas de fenômenos físicos com intervalo de duração inferiores a uma hora. Considerando que o aumento da resolução temporal acarretaria na percepção de fenômenos microescalares, e com isso há uma maior variabilidade do sinal observado, o emprego da técnica desenvolvida neste trabalho tende a ser ainda mais vantajoso quando utilizados dados de resolução temporal intra-horária.

Tendo em vista os resultados preliminares expostos anteriormente, fica claro que a utilização de plantas fotovoltaicas com potências distintas, desde que a contribuição de cada uma das plantas na potência total seja bem estimada, contribui de forma significativa para a melhora da suavização e previsibilidade da potência fotovoltaica gerada.

Como perspectiva futura, podemos salientar desde já a utilização de séries de potência da planta mínima para a estimativa da contribuição de cada uma das centrais na potência total instalada, bem como a posterior comparação com os resultados obtidos, utilizando o pressuposto de que a potência elétrica gerada pelas plantas fotovoltaicas varia linearmente com a irradiância incidente no plano coletor. Essas projeções se colocariam de modo a verificar se existe um ganho apreciável devido à utilização de séries de potência elétrica gerada na estimativa dos coeficientes da regressão linear múltipla.

A exclusiva utilização da metodologia proposta neste trabalho não é suficiente para determinação dos possíveis locais de instalação de centrais, bem como da potência nominal de cada uma dessas centrais. Contudo, ela pode constituir uma ferramenta útil a ser aplicada em estudos mais aprofundados com vistas à escolha de possíveis locais para instalação de centrais fotovoltaicas, fazendo uso de uma base de dados que possua uma maior resolução temporal (intra-horária) e um período de prospecção de dados climatologicamente significativo (ordem de décadas). Esses estudos também devem levar em consideração características econômicas e técnicas relacionadas ao Sistema Interligado Nacional.

REFERÊNCIAS

- ANEEL. 2002. Atlas de Energia Elétrica do Brasil. ANEEL (Agência Nacional de Energia Elétrica). Ministério de Minas e Energia (MME).
- Geladi, P., & Kowalski, B. R. 1986. Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185, 1-17.
- Harper, J. P. 2010. Statistical analysis of solar irradiation in a distributed microgrid. University of California, San Diego.
- Hoicka, C., Rowlands, I. 2011. Solar and wind resource complementarity: Advancing options for renewable electricity integration in Ontario, Canada. *Renewable Energy*, 36, 97-107.
- Hong, T. 2010. Short Term Electric Load Forecasting (a dissertation submitted to the Graduate Faculty of North Carolina State University). North Carolina State University. Raleigh, North Carolina.
- Mills, A., Wisser, R. 2010. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. Ernest Orlando Lawrence Berkeley National Laboratory. Berkley, California.
- Rigollier, C.; Bauer, O.; Wald, L. 2000. On the clear sky model of the ESRA-European Solar Radiation Atlas-with respect to the Heliosat method. *Solar energy*, 68(1), 33-48.
- Thornton, P. E., Running, S. W. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology*, 93(4), 211-228.

STATISTICAL ANALYSIS WITH EMPLOYMENT OF MULTIPLE LINEAR REGRESSION TECHNIQUES FOR VARIABILITY SMOOTHING IN LARGE-SCALE PHOTOVOLTAIC SOLAR GENERATION CONNECTED TO THE BRAZILIAN NATIONAL INTERCONNECTED SYSTEM (SIN)

Abstract. *The intermittent nature of wind and solar power generation, whose participation in the national electricity matrix has increased significantly over the last few years (BEN, 2016), promotes the creation of more sophisticated methods and tools for the operation of the national electricity network. Simultaneously, it is of fundamental importance to consider the installed power spatial distribution, in order to increase the stability of the dispatch of these intermittent resources with a high degree of uncertainty in the National Interconnected System (SIN). In this context, this work presents a simple and efficient method, which makes use of the multiple linear regression model, for the establishment of the nominal power of large-scale photovoltaic plants (power of the order of tens of MW), in order to use complementarity of the solar resource of the several studied sites in order to increase the smoothness and predictability of the electric power generated. In this study, 16 sites of interest were taken into account in the Northeast of Brazil, the region with the highest solar potential available (ANEEL, 2002). The data describing the behavior of the solar resource of each of these locations (on an hourly basis) scale were obtained from the output of physical models provided by the National Solar Radiation Database (NSRDB).*

Key words: *Photovoltaic Solar Energy; Multiple Linear Regression; National Interconnected System*