

UM EXPERIMENTO DE SIMULAÇÃO DE SÉRIES TEMPORAIS DE RADIAÇÃO SOLAR: ANÁLISE EM COMPONENTES PRINCIPAIS DO MODELO SATELITAL GL VERSÃO 1.2

Juan Carlos Ceballos – j.cebal33@gmail.com
Simone Marilene Sievert da Costa – simone.sievert@inpe.br
Divisão de Satélites e Sensores Meteorológicos - DISSM
Coordenação Geral de Ciências da Terra - CGCT
Instituto Nacional de Pesquisas Espaciais - INPE

Resumo. É descrita uma aplicação da Análise em Componentes Principais (ACP) a uma série de 15 anos de médias mensais de irradiância solar (modelo satelital GL1.2, resolução 0.04°, período 2004-2018). O objetivo foi testar a reconstrução de séries temporais em sub-regiões do Brasil, com medições de radiação solar (rede automática do Instituto Nacional de Meteorologia) para comparação. A área geral abrangia desde o Sul até o Nordeste do Brasil, incluindo parte do Cerrado e Amazônia oriental, além de área oceânica vizinha. A estimativa de componentes principais obteve-se a partir da ACP de 132 pixels (sequência de 180 valores). Campos de cargas fatoriais em alta resolução (0.04°) foram construídos como coeficiente de correlação da cada pixel da região com as componentes principais. Apenas 4 modos principais acumulam mais de 80% da variância na região (10 modos acumulam 92%). A regionalização das primeiras quatro cargas fatoriais define claramente regiões de influência, ligadas à Zona de Convergência Intertropical (ZCIT), Zona de Convergência do Atlântico Sul (ZACAS) e Alta Subtropical do Atlântico Sul (ASAS).

Palavras-chave: Energia Solar, Satélite GOES, Séries Temporais

1. INTRODUÇÃO

A correlação entre séries temporais de dados de uma rede de estações de medida permite analisar características comuns de variáveis meteorológicas e sua distribuição espacial. A análise em componentes principais (ACP, no que segue) permite generalizar este conceito: partindo de um conjunto de N estações, cada uma com T informações, com uma matriz de dados $X(N \times T)$ e coeficiente de correlação $Q(N \times N)$, a ACP permite reduzir o conjunto de informações à série temporal de um conjunto M muito menor de “estações” ou “comportamentos principais” com matriz $Z(M \times T)$, capaz de reproduzir o comportamento das N estações de forma simples (linear) e com elevado grau de acurácia. Ao mesmo tempo, o grau de correlação elevado entre a série de dados $X_k(t)$ de algumas estações e os valores $Z_m(t)$ de comportamentos principais permite agrupar (regionalizar) as estações de forma objetiva em M conjuntos característicos.

O presente trabalho aplica a ACP a uma série temporal de irradiância solar média (valores mensais) do modelo GL (GLobal). O modelo é aplicado a imagens do canal VIS-Imager do satélite GOES (Ceballos et al. 2004; Porfirio et al. 2020). Foi escolhido o período 2004-2018, e uma região que se estende desde o sul até o nordeste do Brasil, incluindo o Cerrado e uma faixa oceânica vizinha. Nessa região, foi examinada a distribuição geográfica de modos principais de comportamento da radiação solar e sua relação com fenômenos meteorológicos regionais. Finalmente, foi testada a simulação da série temporal de estações meteorológicas automáticas (rede INMET) em três sub-regiões limitadas.

2. FUNDAMENTOS: APLICAÇÃO DE CONCEITOS DE ACP A UMA REDE SOLARIMÉTRICA

Os elementos teóricos da ACP foram extraídos de Johnson e Wiechern (2007). Seja uma rede solarimétrica com N estações, fornecendo dados $G_k(t)$, $t=1,2,\dots,T$; $k=1,2,\dots,N$. Consideremos os valores padronizados $y_k(t)$ pela média G_{media_k} e desvio padrão S_k

$$y_k(t) = [G_k(t) - G_{media_k}] / S_k, \quad (1)$$

$$G_{media_k} = \sum_t G_k(t) / T \quad (2)$$

$$S_k^2 = \sum_t [G_k(t) - G_{media_k}]^2 / T. \quad (3)$$

A matriz $N \times T$ de dados y_{kt} (k = linha, t = coluna) pode ser imaginada como a evolução temporal de um vetor $Y(t)$ num espaço de N dimensões. A ACP procura enxergar em que medida os vetores Y se alinham ao longo de um conjunto de eixos principais $\{a_1, a_2, \dots, a_m, \dots, a_N\}$. As N componentes principais $z_1(t), z_2(t) \dots z_N(t)$ ou vetores $Z(t)$

descrevem o mesmo sistema de vetores \mathbf{Y} ; a relação entre os vetores \mathbf{Y} e \mathbf{Z} é definida por uma matriz $\mathbf{A}(N \times N)$ de transformação de coordenadas, sendo

$$\mathbf{A} \mathbf{Y}(t) = \mathbf{Z}(t), \quad \mathbf{A}^{-1} = \mathbf{A}', \quad \mathbf{Z}(t) = \mathbf{A}' \mathbf{Y}(t); \quad (4)$$

$$a_{m1} y_{1t} + a_{m2} y_{2t} + \dots + a_{mN} y_{Nt} = z_{mt}, \quad (5)$$

$$a_{1k} z_{1t} + a_{2k} z_{2t} + \dots + a_{Nk} z_{Nt} = y_{kt}. \quad (6)$$

Os vetores \mathbf{a}_m são ortonormais, indicam as direções principais no espaço de N dimensões, e são soluções da equação

$$\mathbf{Q} \mathbf{a} = \lambda \mathbf{a}, \quad (7)$$

onde \mathbf{Q} é a matriz de correlação entre as variáveis y_k , com N autovetores \mathbf{a}_m e autovalores λ_m . Se por um lado as variáveis $y_k(t)$ têm média nula e variância unitária, demonstra-se que as CPs $z_m(t)$ têm média nula, variância λ_m e são mutuamente ortogonais. Assim, as z_{mt} podem ser normalizadas para variáveis w_{mt} , e os coeficientes a_{mk} serem transformados em cargas fatoriais (“weighting factors”) f_{mk} , de forma que a partir da Eq. 6 e da ortonormalidade das variáveis w_{mt} , tem-se

$$w_{mt} = z_{mt} / \sqrt{\lambda_m}, \quad a_{mk} \cdot \sqrt{\lambda_m} = f_{mk}, \quad (8)$$

$$f_{1k} w_{1t} + f_{2k} w_{2t} + \dots + f_{Nk} w_{Nt} = y_{kt} \quad (9)$$

$$f_{1k}^2 + f_{2k}^2 + \dots + f_{mk}^2 + \dots + f_{Nk}^2 = 1 \quad (10)$$

As equações (9) e (10) são fundamentais. Por um lado, a Eq. 9 representa uma descrição *exata* da k -ésima variável; por outro lado, na Eq. 10 os termos f_{mk}^2 representam a contribuição da m -ésima CP para a variância da k -ésima estação e dependem fortemente da grandeza dos autovalores λ_m . Apenas algumas CPs (sejam $M \ll N$) poderiam acumular uma elevada fração de sua variância, não se justificando um desenvolvimento maior (o resto seria, muito provavelmente, constituído por “erros” ou flutuações que não contribuem à informação sobre y_{kt}). Finalmente, se o conjunto de estações abrange uma área geográfica extensa, observar a distribuição de valores do fator f_{m*} permite regionalizar a influência da m -ésima componente principal sobre a radiação solar.

3. MÉTODO DE ANÁLISE

Valores diários do modelo GL são estimados para América do Sul e região oceânica vizinha, fornecendo informação em 1800×1800 pixels, resolução espacial de $0,04^\circ$. Foi considerada a série temporal dos valores mensais no período 2004-2018 (180 informações por pixel). Escolheu-se a região R: $[|-60 < \text{LON} < -31^\circ, -30 < \text{LAT} < 0^\circ]$, contendo 751 linhas \times 726 colunas = 545226 pixels. A Fig. 1 ilustra a distribuição geográfica da média GLmedia e do desvio padrão para cada pixel. Para evitar uma matriz de correlação \mathbf{Q} com dimensão intratável, foram escolhidos 132 pixels num padrão geográfico regular, como ilustrado na Fig. 1. Programas em MatLab (compatíveis Octave) permitiram resolver a Eq. 7, determinando as matrizes \mathbf{A} e \mathbf{Z} .

A matriz \mathbf{A} permitiria apenas uma visão grosseira da distribuição geográfica de cargas fatoriais f_{m*} na região R. Considerando que f_{mk} representa o coeficiente de correlação entre a variável $y_k(t)$ original e a $w_m(t)$ principal, e que esta é igualmente válida para todos os pixels, foi calculado o coeficiente de correlação para todos os pixels em R. Gerou-se assim campos de cargas fatoriais em alta resolução.

Foram escolhidas três regiões restritas, contendo estações da rede automática do INMET em 2017: uma no Sul, uma no Cerrado, e outra no Nordeste. Para esse ano e esses locais, foi realizada a simulação da série de GL segundo as Eqs. 1 e 9, mas restringindo o desenvolvimento aos primeiros 10 modos principais (dentre 132).

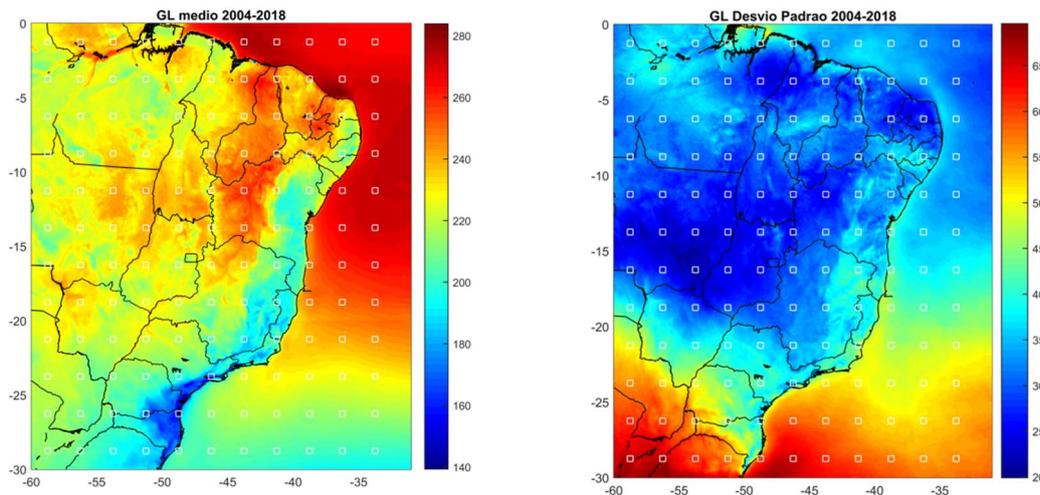


Figura 1. Irradiância solar no Sul, Sudeste, Centro Oeste e Nordeste do Brasil, período 2004-2018. Baseada na série temporal de médias mensais de GL: valor médio e desvio padrão da série. Unidades: $W.m^{-2}$ (equivalência: $100 W m^{-2}$ médios \leftrightarrow $8,64 MJ m^{-2}$ diários \leftrightarrow $2,4 kWh m^{-2}$ diários).

4. RESULTADOS

A Fig. 1 mostra que a irradiância solar média anual sobre o continente apresenta basicamente três regiões distintas. Valores mínimos de irradiância ($GL \leq 200 \pm 40 W m^{-2}$) são encontrados sobre a região litorânea do Brasil, onde o regime de nebulosidade é influenciado pelas massas de ar tropical marítima que transportam umidade para o continente ao longo de todo o ano. Máximos de radiação solar ($GL > 240 \pm 30 W m^{-2}$) observam-se sobre as regiões de semiárido, onde a atuação das massas de ar é irregular. Sobre o Centro Oeste, Sudeste e parte da Amazônia Oriental verificam-se valores de irradiância em torno de 200 a 230 $W m^{-2}$, onde diversas escalas globais e regionais de tempo influenciam a nebulosidade local, e portanto o regime de radiação. Sobre o Oceano Atlântico Tropical observam-se valores acima de 250 $W m^{-2}$, e na parte subtropical valores abaixo de 220 $W m^{-2}$.

A Fig. 2 é um resumo da análise fatorial para a região R, período 2014-2018. Os autovalores decrescem rapidamente, atingindo valor unitário no oitavo modo e acumulando 91,5% da variância. Para o modo 10, $\lambda_{10} = 0,67$ e a variância acumulada é de 92,6%. Na mesma Figura percebe-se que o cálculo de coeficiente de correlação com as componentes principais permite uma descrição coerente e detalhada da influência dos modos principais. Na primeira figura à direita é representada a variância acumulada pelos 4 primeiros modos. Observa-se acumulação superior a 80%, exceto em regiões restritas, algumas delas no Mato Grosso, Rio Grande do Norte + Paraíba + setor de Pernambuco, interior da Bahia. Contudo, a variância acumulada nessas sub-regiões é 75%. Conclui-se que os quatro primeiros modos já devem descrever características de extensas áreas da região R. Com efeito,

- O modo 1 concentra-se a área oriental e no oceano. A isolinha $f_1 = 0,7$ demarca a cobertura de 50% da variância local, e $f_1 = 0,9$ (81% da variância) é associado a esse modo no oceano, não longe do litoral do Brasil.
- O modo 2 substitui o modo 1 no interior do Nordeste, incluindo Tocantins e norte do Mato Grosso. Isolinha $f_2 = 0,9$ é observada principalmente numa região pré-amazônica.
- Os modos 3 e 4 ocupam uma área ao norte e ao sul, respectivamente, da divisa entre São Paulo + Mato Grosso do Sul e Minas Gerais + Goiás. Eles existem numa região parcialmente não influenciada pelos modos 1 e 2. Contudo, o peso do fator f_3 não excede 40% da variância local (numa sub-região vizinha a Brasília).

A Figura 3 apresenta a evolução temporal das três primeiras componentes principais $w(t)$ ao longo de 15 anos. É importante notar que elas não representam regiões específicas, como não seja através das cargas fatoriais f . Observa-se que todas elas têm forte componente anual (evidenciada pela presença de 15 máximos na série), mas não necessariamente comportamento regular. Alguns eventos de tipo meteorológico/climático podem ser identificados na combinação das variáveis temporais $w_m(t)$ e os parâmetros espaciais f_m .

A primeira componente principal tem máximos durante o início da primavera até o início do outono (Setembro à Março), e mínimos nos meses de inverno (Junho e Julho), indicando que parte da variabilidade da radiação, como esperado, está associada à declinação dos raios solares, sobretudo nas regiões onde o fator 1 apresenta alta correlação. Adicionalmente, as componentes principais 1 e 2 indicam que a variabilidade da radiação solar está relacionada aos regimes atmosféricos de sistemas de monção que atuam durante o período de verão, principalmente a Zona de Convergência Intertropical (ZCIT) e Zona de Convergência do Atlântico Sul (ZCAS).

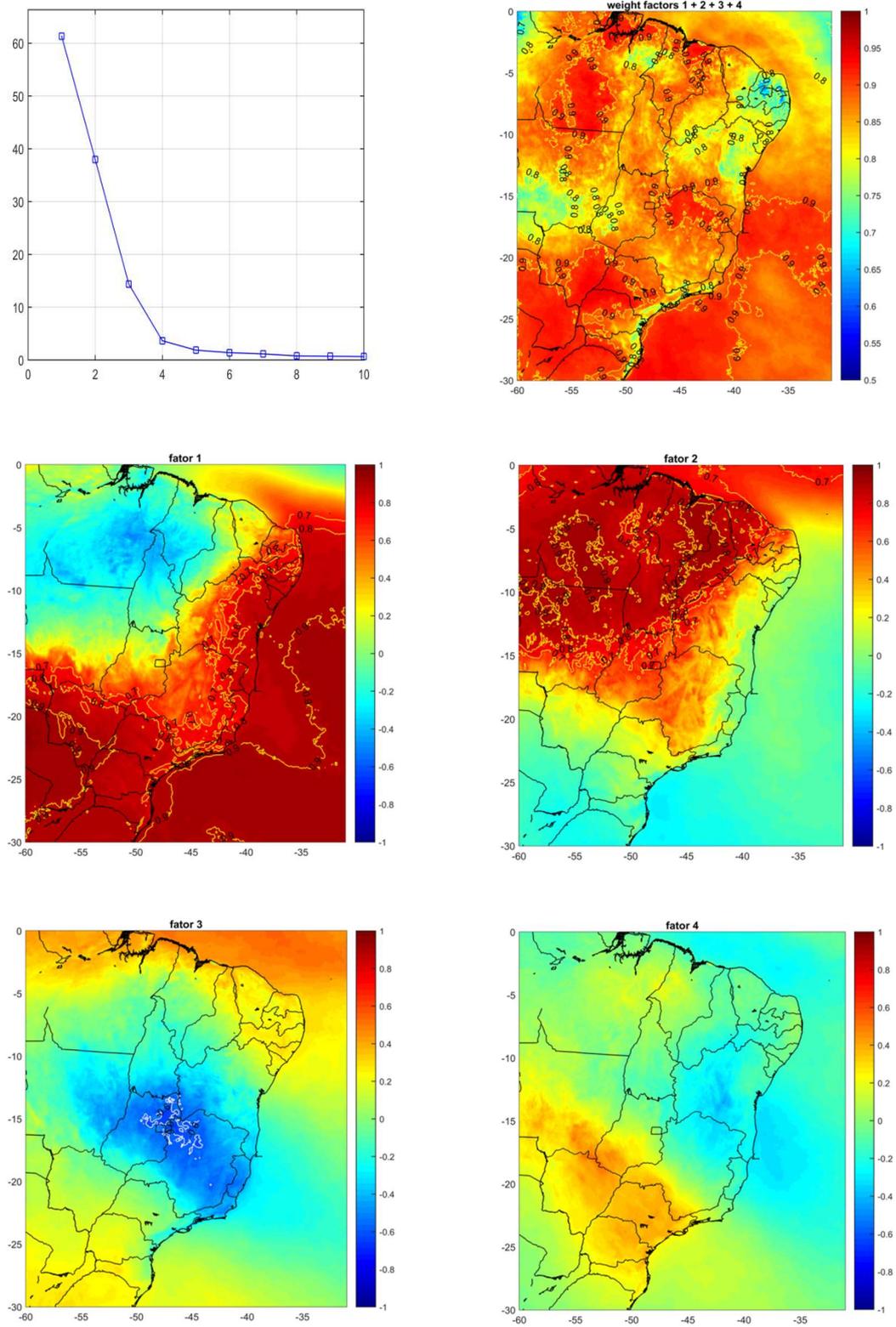


Figura 2. Autovetores, contribuição das primeiras 4 cargas fatoriais, associação regional com cada modo. Unidades: adimensionais.

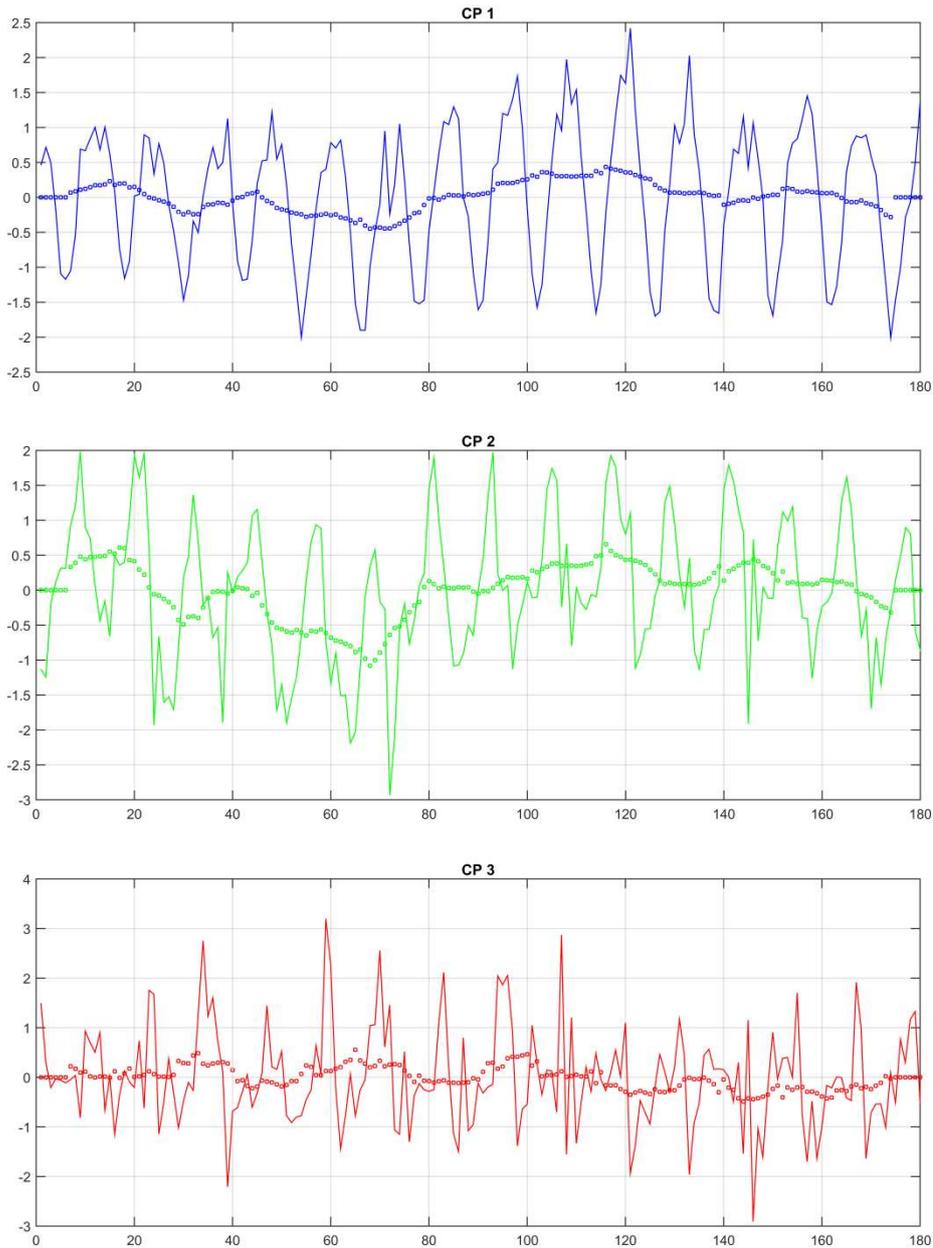


Figura 3. Primeiras três CP's (adimensionais) na Região 1. Está incluída a evolução da média móvel (largura: 12 valores, de t-6 a t+5).

Os fatores 3 e 4 também apresentam valores significativos de correlação (acima de 0,6), em uma área menor. Os dois fatores juntos formam um dipolo, sendo que o segundo tem um sinal positivo ao sul, enquanto que o fator 3 apresenta correlação negativa ao norte com máximos da componente principal 3 nos meses de julho, agosto e setembro. Essas duas componentes principais indicam que efeitos sinóticos, como sistemas frontais, influenciam os modos de variabilidade da radiação nessa região. Na época de inverno, algumas frentes podem avançar sobre a região sudeste e sul da Bahia, ou ainda ficar estacionárias devido a atuação da Alta Subtropical do Atlântico Sul (ASAS), de forma que esta também influencia nos campos de radiação solar. Uma análise mais aprofundada da relação entre sistemas de larga escala/sinóticos e o regime local de radiação solar está além do escopo deste trabalho. Futuras análises podem ser aprimoradas utilizando a ACP em áreas específicas, e métodos de separação de modos como a rotação Varimax (Johnson e Wiechern, 2007).

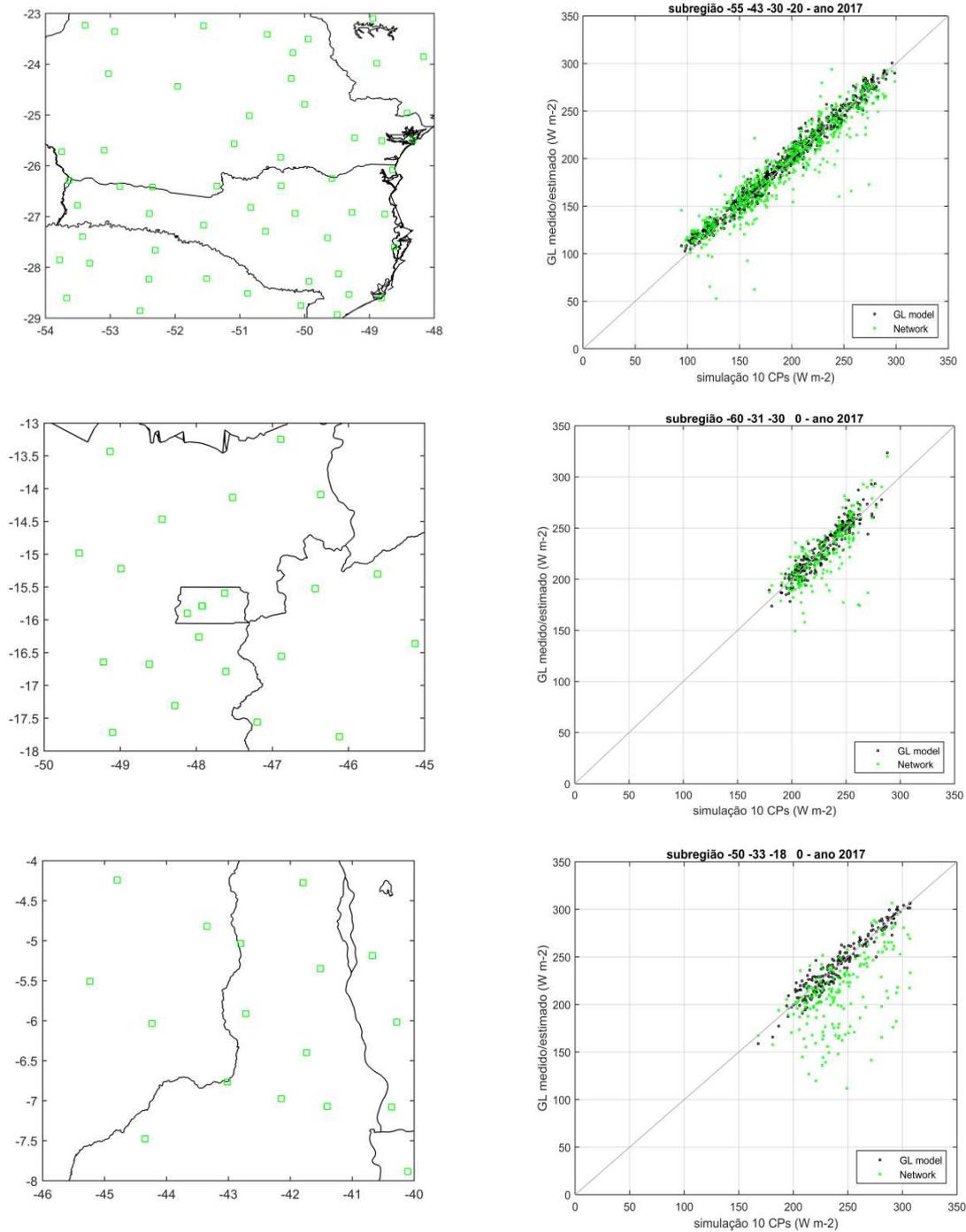


Figura 4. Simulação de um período anual (2017) de médias mensais a partir das 10 primeiras CPs. Aplicação a três subregiões: Sul (Paraná/Santa Catarina), Cerrado (Brasília), Nordeste semiárido (Piauí/Maranhão). Comparação com dados de sites do INMET (verde) e os correspondentes no GL 1.2 (pretos). Localização indicada nos mapas à esquerda.

A Figura 4 ilustra o resultado da simulação de séries temporais em 3 sub-regiões, utilizando uma aproximação de ordem 10 (primeiros 10 modos principais). Foi escolhido o ano de 2017 para simular, e locais com existência de estações automáticas da rede do INMET. Como esperado, a simulação representa de perto os valores do GL 1.2. (existe portanto alta coerência entre as características da região e sua descrição em componentes principais); 10 CPs se apresentam como um número suficiente de modos a considerar. Por outro lado, fica evidenciada a qualidade da simulação dos valores medidos pela rede meteorológica (embora, no caso de Nordeste, um número elevado de estações tinha dados anômalos ou faltosos). Um resultado colateral é que os dados do modelo GL e das estações se correspondem estreitamente, evidenciando a qualidade do modelo.

5. CONCLUSÕES

O experimento descrito, aplicado a uma extensa área do Brasil incluída na base de dados do modelo GL1.2, mostrou a qualidade e a simplicidade da ACP como ferramenta de diagnóstico para regionalizar comportamentos e colocar em evidência regimes de radiação solar (influenciados, no caso, pela oscilação anual da declinação solar, e pela ZCIT, ZACAS e ASAS). Estudos futuros permitirão aprofundar na descrição de características espaciais e temporais da radiação solar na América do Sul e sua dependência de sistemas sinóticos e de larga escala, restringindo convenientemente as áreas de estudo.

Ficou evidente que os 10 primeiros modos principais acumulam mais de 90% da informação das séries temporais. Dessa forma, a ACP de 15 anos de dados pode ser realizada com, no máximo, uma matriz de componentes principais com dimensão 10 modos×180 meses, e 10 matrizes de cargas fatoriais em alta resolução (0,04°). A aplicação a regiões mais restritas (que amplificariam a definição das relações sistemas atmosféricos × radiação solar), requer de arquivos digitais ainda menores.

Os resultados do experimento de simulação, aplicado à série anual de um conjunto de sítios com medições de radiação solar, mostram que o ajuste com 10 modos principais produz um ajuste excelente: (a) com os dados originais de GL; (b) com os dados de medição. Além de produzir informação sobre séries temporais em locais sem dados, o experimento evidenciou também o bom ajuste entre dados de GL e medições de rede solarimétrica.

6. REFERÊNCIAS

- Ceballos, J.C., M.J. Bottino, J.M. Souza, 2004. A simplified physical model for assessing solar radiation over Brazil using GOES 8 visible imagery. *J. Geophys. Research* vol. 109, D02211, doi:10.1029/2003JD003531.
- Johnson, R. A. and Wichern, D. N., 2007. *Applied Multivariate Statistical Analysis*, 6th edition, pp. 773 (Prentice Hall).
- Porfirio, A.C.S., J.C. Ceballos, J.M.S. Britto, S.M.S. da Costa, 2020. Evaluation of Global Solar Irradiance Estimates from GL1.2 Satellite-Based Model over Brazil Using an Extended Radiometric Network. *Remote Sensing* 2020, 12, 1331; doi:10.3390/rs12081331

A NUMERICAL EXPERIMENT OF SIMULATION OF SOLAR RADIATION TIME SERIES; PRINCIPAL COMPONENTS ANALYSIS OF SATELLITE-BASED MODEL GL 1.2

Abstract. *It is described na application of Principal Components Analysis (PCA) to a 15 years series of monthly values of mean solar irradiance (satellite-based model GL 1.2, resolution 0.04°). It was aimed to test building time series in some Brazilian sub-regions with solarimetric sites of National Institute of Meteorology, in order to have reference ground-based values for comparison. The general study area was extended from Southern to Northeastern Brazil, including part of Cerrado, East Amazonia and neighboring oceanic area. PCs were estimated from only 132 pixels time series (180 months). High resolution factor loading fields were obtained as correlation coefficient of pixel and PC time series. The first for principal modes cumulate more than 80% of regional variance (10 modes summed 92%). Geographical distributions of factor loadings make evident the influence of Intertropical Convergence Zone (ITCZ), South Atlantic Convergence Zone (ZACAS) and South Atlantic Subtropical High. Time series of GL, as well as those of INMET network, are well described by the first 10 principal modes development. Testing subregions were located in South, Cerrado and inner Northeast.*

Keywords: Solar energy, GOES satellite, Time series.