



## IMPUTAÇÃO DE DADOS DIÁRIOS DE RADIAÇÃO SOLAR GLOBAL VIA AMBIENTE R

Danilo Siden Batista Santos Silva<sup>1</sup>, Juan Carlos Vilalta Jimenez<sup>2</sup>, Charla Campos Pereira<sup>3</sup>, Givanildo de Gois<sup>4</sup>, José Francisco de Oliveira Júnior<sup>5</sup>

1. Estudante de Graduação em Meteorologia da Universidade Federal de Alagoas (UFAL), Maceió/AL – Brasil
2. Estudante de Graduação em Meteorologia da Universidade Federal de Alagoas (UFAL), Maceió/AL – Brasil
3. Estudante de Graduação em Meteorologia da Universidade Federal de Alagoas (UFAL), Maceió/AL – Brasil
4. Pós-Doutorando da Escola de Engenharia Industrial Metalúrgica de Volta Redonda, Universidade Federal Fluminense (UFF), Volta Redonda, RJ - Brasil
5. Professor Doutor do Instituto de Ciências Atmosféricas (ICAT) da Universidade Federal de Alagoas (UFAL), Maceió/AL – Brasil. E-mail: jose.junior@icat.ufal.br

Recebido em: 06/04/2019 – Aprovado em: 10/06/2019 – Publicado em: 30/06/2019  
DOI: 10.18677/EnciBio\_2019A76

### RESUMO

O objetivo do estudo foi avaliar a técnica de imputação de dados via *software* Ambiente R na série temporal diária de radiação solar global ( $Q_g$ ) no município de Arapiraca, Alagoas - Nordeste Brasileiro (NEB). A série temporal de radiação solar global de 2008 a 2016 foi obtida do INMET. A partir dos percentuais de falhas na faixa entre 0,05% e 0,15%, foi aplicado o método de imputação múltipla aos dados de  $Q_g$ , via pacote MICE. O teste de PBIAS apontou superestimava da radiação solar global observada ( $R_0$ ) em Arapiraca (-7,40). Valores altos de NSE (0,90) e  $R^2$  (0,99) também foram identificados na estação Arapiraca. Isso mostrou um desempenho muito bom do pacote MICE entre  $R_0$  e a radiação solar global imputada ( $R_i$ ). Os diagramas de dispersão da  $R_0$  e  $R_i$  mostrou correlação ( $r$ ) quase perfeita e os índices de concordância ( $d$ ) e confiança ( $c$ ) categorizados como ótimo e muito bom, em Arapiraca ( $r = 0,99$  e  $d = 0,97$ ). Os parâmetros EPE, RMSE e MAE indicaram uma baixa dispersão do modelo de regressão linear simples em Arapiraca EPE (113  $\text{KJ.m}^{-2}$ ), RMSE (137,49  $\text{KJ.m}^{-2}$ ) e MAE (111,47  $\text{KJ.m}^{-2}$ ). O teste t de Student aplicado à série temporal de  $Q_g$  mostrou que a  $H_0$  foi rejeitada para p. valor < 0,05.

**PALAVRAS-CHAVE:** Preenchimento de Falhas, Série Temporal, Radiação Solar Incidente, *Software* Livre.

# IMPUTATION OF THE DAILY TIME SERIES OF GLOBAL SOLAR RADIATION BY ENVIRONMENT R

## ABSTRACT

The purpose of this study was to evaluate the data imputation technique through R software environment with daily time series of global solar radiation ( $Q_g$ ) in the county of Arapiraca, Alagoas - Brazilian Northeast (NEB). The time series of global radiation from 2008 to 2016 was obtained by INMET. From the percentage of gaps in the range between 0.05% and 0.15% it was applied the multiple imputation method to the  $Q_g$  data by MICE pack. The PBIAS test indicated an overestimated observed global solar radiation ( $R_0$ ) in Arapiraca (-7.40). High values of NSE (0.90) and  $R^2$  (0.99) were identified in the Arapiraca station. This demonstrated a very good performance of the MICE pack between  $R_0$  and imputed global solar radiation ( $R_i$ ). The scatter plots of  $R_0$  and  $R_i$  showed an almost perfect correlation ( $r$ ) and the Indexes of agreement ( $d$ ) and trust ( $c$ ) categorized with great and very good in Arapiraca ( $r=0.99$  and  $d=0.97$ ). The EPE, RMSE and MAE parameters indicated a low dispersion of the simple linear regression model in Arapiraca EPE ( $113 \text{ KJ.m}^{-2}$ ), RMSE ( $137.49 \text{ KJ.m}^{-2}$ ) and MAE ( $111.47 \text{ KJ.m}^{-2}$ ). The t-student test applied in the time series of  $Q_g$  showed that  $H_0$  was rejected for  $p$ . value  $< 0.05$ .

**KEYWORDS:** Gap Filling, Time Series, Incident Solar Radiation, Free Software.

## INTRODUÇÃO

A radiação solar global ( $Q_g$ ) é a radiação solar que atinge a superfície do solo, sendo formada pelas componentes direta e difusa (LYRA et al., 2015). O conhecimento das variações sazonais e temporais de  $Q_g$  permitem sua utilização em estudos de evapotranspiração - ET (GOIS et al., 2016), otimização de demanda em irrigação, previsão de safras, mudança climática (TEODORO et al., 2018), desenvolvimento de sistemas tecnológicos de energia (térmicos e fotovoltaicos), entre outros (SOUZA et al., 2017).

Um dos graves problemas existentes no Brasil é a ausência de informação de dados climáticos, principalmente dados de  $Q_g$ , em virtude de falhas operacionais ou que os dados de  $Q_g$  estão parcialmente ausentes consecutivamente por dias (GOIS et al., 2016; LAYANUN et al., 2017). Recentemente, têm sido utilizados métodos de imputação, por exemplo, os que utilizam a decomposição por valores singulares de uma matriz (DVS) – (GARCÍA-PEÑA et al., 2014). Vale destacar que o problema com dados faltantes (*missing data*) em séries temporais, não se restringe apenas as áreas de Meteorologia e Climatologia (ZHANG, 2016; KOWARIK ; TEMPL, 2016, GRÜND et al., 2016), se estende também as áreas da Saúde e das Ciências Sociais (NUNES et al., 2010).

Além dos problemas com dados de  $Q_g$ , o uso de dados hidrometeorológicos também é comprometido, principalmente no Brasil. Os dados hidrometeorológicos são essenciais para o planejamento e gerenciamento dos recursos hídricos, principalmente com relação à drenagem urbana, abastecimento d'água e avaliação de seca prolongada (VIEIRA et al., 2018). Estes dados, geralmente são organizados em séries temporais, que podem estimar índices, no caso índices de seca (TERASSI et al., 2018; OLIVEIRA JÚNIOR et al., 2018) e, posteriormente indicadores de tendências (SOBRAL et al., 2019), nos quais os gestores se baseiam nestas ferramentas para tomada de decisão. Porém, a base de dados meteorológicos provenientes das estações automáticas ou

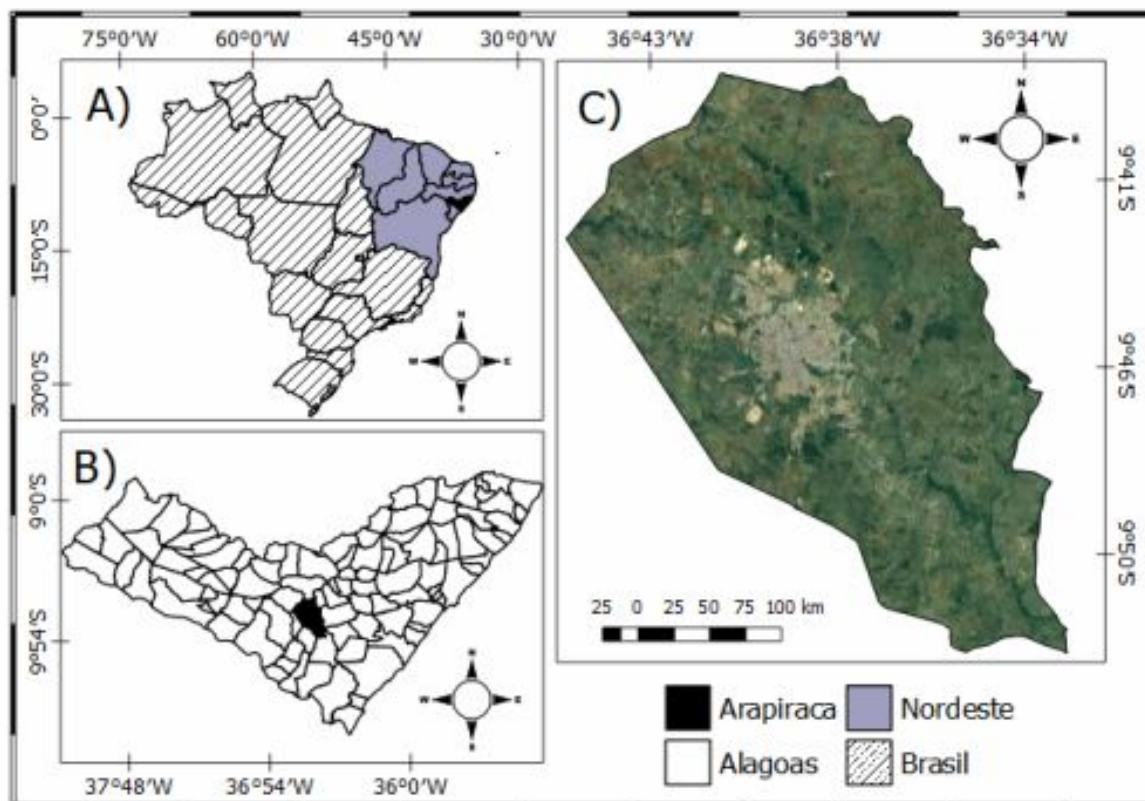
convencionais do INMET, juntamente com a Hidroweb da Agência Nacional de Águas (ANA) entre outras bases de dados do Governo são falhas e, portanto necessitam de técnicas de preenchimento de falhas para consolidar uma série temporal consistente (BRITO et al., 2016; LYRA et al., 2017).

Baseado no exposto, objetivo desse estudo é avaliar a técnica de imputação de dados via *software* Ambiente R na série temporal diária de radiação solar global no município de Arapiraca, Alagoas - Nordeste Brasileiro (NEB).

## MATERIAL E MÉTODOS

### Área de estudo

Arapiraca é um município na mesorregião do Agreste de Alagoas com um território de 345.655 km<sup>2</sup> e uma altitude média de 264 m. Segundo o IBGE (2018) a população estimada de Arapiraca é de 234.185 pessoas em 2017, isto é, com a 2ª maior população do Estado. Ela possui clima tipo “C1A's2w” caracterizado por ser subúmido, megatérmico, com grande deficiência de água no verão, e excesso de água moderada no inverno (SANTOS, 2012). O milho e o fumageiro são os principais cultivo do município.



**FIGURA 1.** Localização do Nordeste Brasileiro (A), do Estado de Alagoas (B) e do Município de Arapiraca (C).

### Série temporal e organização dos dados

Foram utilizados dados horários de radiação solar global ( $Q_g$ , KJ.m<sup>-2</sup>) da estação meteorológica automática (EMA) do município de Arapiraca, disponíveis no Instituto Nacional de Meteorologia (INMET) (<http://www.inmet.gov.br/>) no período de 2008 a 2016. Para a

organização e cálculos dos dados foi usado *software* ambiente R versão 3.4.3 (R DEVELOPMENT CORE TEAM, 2017).

Os dados foram organizados cronologicamente a partir da união de duas séries temporais: i) de abril de 2008 até 2014 e ii) início de 2015 até o final de 2016. A partir disso, foram calculados os totais acumulados diários, mensais e anuais. Em ambas as séries temporais havia diversas falhas, sendo contabilizado o total e o percentual (%) diários para os horários das 9 horas às 22 horas, através, da aplicação da função “*mstats*” via *software* R versão 3.4.3 (R DEVELOPMENT CORE TEAM, 2017). Essa função tem por objetivo calcula a proporção de observações ausentes “*missing*” em um determinado conjunto de dados por linhas e colunas, conforme os comandos a seguir (1), (2), (3) e (4):

```
mstats(dados1) (1)
```

```
FALHAS=c(mstats(dados1)) (2)
```

```
(TAB_FALHAS_1=FALHAS$columns) (3)
```

```
write.xlsx (TAB_FALHAS_1,"MISSING_2008-2016-DIARIO.xlsx", row.names = T) (4)
```

Em que, dados1= planilha de dados de radiação solar global (KJ.m-2); mstats = função para calculo de dados faltoso no R; TAB\_FALHAS\_1 = tabela de dados faltosos por coluna; write.xlsx = saída de dados com extensão para o Excell.

### **Escolha do Método de Imputação**

A escolha do melhor método de imputação de dados, neste estudo baseia-se na técnica sugerida por Harrel (2001), em que os percentuais de dados faltantes da variável sobre análise devem seguir os seguintes parâmetros:

- Percentual 0,05: Aplica-se a imputação única ou analisar somente os dados completos;
- Percentual entre 0,05 e 0,15: Imputação única pode ser usada aqui provavelmente sem problemas, entretanto o uso da imputação múltipla é indicado;
- Percentual 0,15: A imputação múltipla é indicada na maior parte dos casos.

Vale ressaltar que se houverem diversos preditores com dados faltantes devem ser feitas as mesmas considerações acima, mas os efeitos da imputação de dados serão mais pronunciados.

Baseado na metodologia de Harrel (2001) foi aplicado o processo de imputação múltipla, pelo pacote “MICE”. O pacote MICE realiza a imputação múltipla baseado na Especificação Condicional Total (ECT), segundo Buuren e Oudshoorn (2011). Vale ressaltar que cada variável possui seu próprio modelo de imputação. Modelos de imputação internos são fornecidos para dados contínuos (correspondência média preditiva, normal), dados binários

(regressão logística), dados categóricos não ordenados (regressão logística politômica) e dados categóricos ordenados (odds proporcionais).

O MICE, também pode imputar dados contínuos de dois níveis (modelo normal, pan, variáveis de segundo nível). Assim, neste estudo foi utilizado o método “midastouch” (correspondência de média preditiva ponderada), para dados contínuos de radiação solar global diário em 24 horas, para os respectivos horários de 9 horas às 22 horas, de acordo com as Eqs. (5), (6), (7), (8) e (9), a seguir:

```
library(mice) (5)
```

```
imp1 <- mice(dados1, method = "midastouch", m=3, maxit = 5) (6)
```

```
dslist1 <- complete(imp1, c(3,5), mild = TRUE) (7)
```

```
(P3=data.frame(dslist1$'3')) (8)
```

```
write.xlsx(P3,"DADOS_3_ARAPIRACA-1.xlsx",col.names=TRUE) (9)
```

Em que, library (mice) = biblioteca do pacote MICE; dados1 = base de dados de radiação solar global diária em 24 horas; Method= método de imputação da variável radiação solar diária em 24 horas; m = número de interações para a variável radiação solar diária em 24 horários; maxit= número de interações máximas para a variável radiação solar diária em 24 horas; dslist1 = função empírica para variável radiação solar diária em 24 horas; P3 = tabela de dados imputados; write.xlsx = saída de dados com extensão para o Excel. Maiores detalhes podem ser verificados no site: <https://cran.r-project.org/web/packages/mice/mice.pdf>.

### Análise estatística

Para avaliar o desempenho dos dados de  $Q_g$ , observados ( $R_o$ ) e imputados ( $R_i$ ), foram usados os seguintes parâmetros estatísticos: os coeficientes de determinação ( $R^2$ ) e correlação ( $r$ ) de Pearson, Índice de Concordância ( $d$ ) de Wilmott (WILLMOTT, 1981), índice de confiança ou desempenho ( $c$ ) de Camargo e Sentelhas (1997), Erro Padrão da Estimativa (EPE,  $KJ.m^{-2}$ ), Raiz Quadrada do Erro Quadrático Médio (RMSE,  $KJ.m^{-2}$ ), Erro Absoluto Médio (MAE,  $KJ.m^{-2}$ ), coeficiente de eficiência Nash-Sutcliffe (NSE) e Percent Bias (PBIAS) Moriasi et al. (2007), teste  $t$  de Student.

O uso desses indicadores é possível identificar os ajustes de superestimativa ou subestimativa geral da amostra, todos esses procedimentos foram realizados via *software* ambiente R versão 3.4.3 (R DEVELOPMENT CORE TEAM, 2017), conforme as Equações 10, 11, 12, 13, 14, 15, 16 e 17 listadas a seguir:

$$R^2 = 1 - \frac{\sum_{i=1}^n \left( R_i - \bar{R}_o \right)^2}{\sum_{i=1}^n \left( R_o - \bar{R}_o \right)^2} \quad (10)$$

$$R = \frac{\sum_{i=1}^n (R_o - \bar{R}_o)(R_i - \bar{R}_i)}{\sqrt{\sum_{i=1}^n (R_o - \bar{R}_o)^2 - \sum_{i=1}^n (R_i - \bar{R}_i)^2}} \quad (11)$$

Os valores de coeficientes de correlação foram classificados segundo a metodologia de Hopkins (2009), conforme a Tabela 1.

**TABELA 1.** Coeficiente de correlação e a sua respectiva classificação.

Coeficiente de Correlação (r)	Classificação
0 – 0,1	Muito baixa
0,1 – 0,3	Baixa
0,3 – 0,5	Moderada
0,5 – 0,7	Alta
0,7 – 0,9	Muito alta
0,9 – 1,0	Quase perfeita

$$d = 1 - \left[ \frac{\sum_{i=1}^n (R_i - R_o)^2}{\sum_{i=1}^n \left( |R_i - R_o| + |R_o - \bar{R}_o| \right)^2} \right] \quad (12)$$

Em que,  $R^2$  = coeficiente de determinação,  $r$  = coeficiente de correlação;  $d$  = índice de concordância de Willmott;  $n$  = número de observação;  $R_i$  é o valor  $i$ -ésimo da radiação solar global imputada; e  $R_o$  é o valor  $i$ -ésimo da radiação solar global observa.

Analogamente, foi empregado o índice confiança ou desempenho ( $c$ ), proposto por Camargo e Sentelha (1997), obtido pelo produto dos índices de correlação ( $r$ ) e concordância ( $d$ ), dados pela Eq. 13 e classificado segundo a Tabela 2:

$$c = r \times d \quad (13)$$

**TABELA 2.** Índice  $c$  proposto por Camargo e Sentelha (1997).

Valor de $c$	Desempenho
> 0,80	Ótimo
0,76 – 0,85	Muito Bom
0,66 – 0,75	Bom
0,61 – 0,65	Mediano
0,51 – 0,60	Sufrível
0,41 – 0,50	Mau
0,40	Péssimo

$$EPE = \sqrt{\frac{\sum_{i=1}^n (R_o - R_i)^2}{n-2}} \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (R_o - R_i)^2}{n}} \quad (15)$$

A seguir, são apresentadas as Eqs. (16 e 17) para a determinação dos valores do NSE e PBIAS, respectivamente.

$$NSE = 1 - \frac{\sum_{i=1}^n (R_o - R_i)^2}{\sum_{i=1}^n (R_o - \bar{R}_o)^2} \quad (16)$$

$$Pbias = 1 - \frac{\sum_{i=1}^n (R_o - R_i)}{\sum_{i=1}^n (R_o)} \times 100 \quad (17)$$

em que n= Número de observações durante o período simulado;  $R_o$  = Valor i-ésimo dados de radiação solar global observado e  $R_i$  = Valor i-ésimo dos dados de radiação solar global imputado,  $\bar{R}_o$  = Média aritmética dos dados de radiação solar global observados.

Viola et al. (2009) e Baltpkpski et al. (2010) adotaram critério de que, quando o valor de NSE for maior que 0,75, o desempenho do modelo é considerado bom. Para valores de NSE entre 0,36 e 0,75, o desempenho é considerado aceitável, enquanto valores de NSE inferiores a 0,36 fazem com que o modelo seja julgado inaceitável. Entretanto, Moriasi et al. (2007) acrescentam os valores de Pbias e  $R^2$  na sua avaliação, além de outros intervalos de classe referentes aos valores de NSE. Por isso, neste estudo a avaliação do desempenho do modelo será realizada conforme a metodologia de Moriasi et al. (2007) - (Tabela 3).

**TABELA 3.** Classificação de desempenho em função de intervalos de classe de NSE, Pbias e  $R^2$ .

NSE	Pbias	$R^2$	Classificação do desempenho
0,75 < NSE 1,00	PBIAS ±10	0,75 < $R^2$ 1,00	Muito Bom
0,60 < NSE 0,75	10 < PBIAS ±15	0,60 < $R^2$ 0,75	Bom
0,50 < NSE 0,60	15 < PBIAS ±25	0,50 < $R^2$ 0,60	Satisfatório
NSE 0,50	PBIAS ±25	$R^2$ 0,50	Insatisfatório

Fonte: Adaptado de MORIASI et al. (2007).

E por fim foi aplicado o teste t de *Student* (Eq. 18), aos dados observados e imputados pelo pacote MICE, para avaliar se existem

diferenças entre as médias dos dados observados e imputados. O teste t é um teste de hipótese que usa conceitos estatísticos para rejeitar ou não uma hipótese nula quando a estatística de teste segue uma distribuição t de *Student*, Normando et al. (2010). Foram formuladas as hipóteses  $H_0$  e  $H_1$ .

$H_0: \bar{R}_o = \bar{R}_i$  (A média dos dados de radiação solar global, observados e imputados não se difere). Para p.valor > 0,05 aceita-se  $H_0$  (S);

$H_1: \bar{R}_o \neq \bar{R}_i$  (A média dos dados de radiação solar global, observados e imputados se difere). Para p.valor < 0,05 rejeita-se  $H_0$  (NS).

$$t = \frac{(\bar{R}_o - \bar{R}_i)}{\sqrt{\frac{S_o^2}{n_o} + \frac{S_i^2}{n_i}}} \quad (18)$$

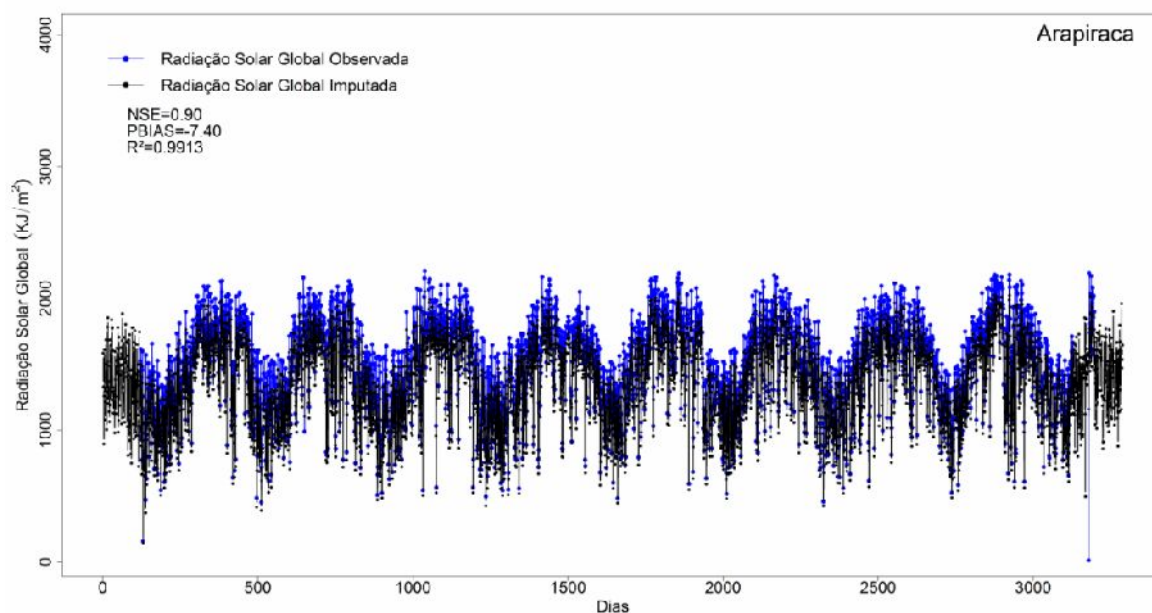
Em que  $t$  = teste t de Student;  $\bar{R}_o$  e  $\bar{R}_i$  = Média aritmética dos dados de radiação solar global, observados e imputados;  $S_0$  e  $S_1$  = Desvios Padrão dos dados de radiação solar global, observados e imputados;  $n_0$  e  $n_1$  = Número de observações durante o período simulado.

## RESULTADOS E DISCUSSÃO

A série temporal de dados  $R_0$  e a reconstruída com dados  $R_i$  na escala diária mostrou que o método MICE consegue mostrar a variabilidade temporal dos valores  $R_i$ . Vale ressaltar que o tamanho do banco de dados afeta os resultados obtidos, desde as calibrações e/ou análises de desempenho, segundo Souza et al. (2017). O coeficiente  $R^2$  foi próximo de 1 (0,9913), categorizado como quase perfeita, conforme a Tabela 1. Vale destacar que o coeficiente  $R^2$  é um índice de ajustamento por regressão linear, que descreve a proporção da variância total dos dados observados pelos dados estimados entre 0 a 1, com valores próximo de 1 com melhor concordância entre os dados observados e estimados (SOUZA et al., 2017), esse mesmo procedimento é utilizado em modelos empíricos de  $Q_g$  e em estudos de cenários de mudanças climáticas (LYRA et al., 2015; GOIS et al., 2016; TEODORO et al., 2018).

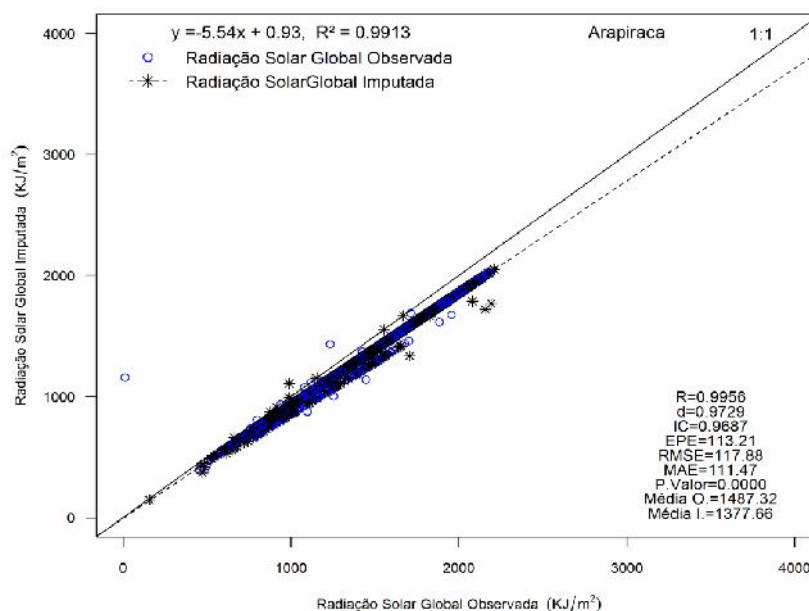
Baseado na Tabela 3, o índice NSE e o coeficiente  $R^2$  foram classificados como muito bom, ao contrário, do índice PBIAS que foi categorizado como bom. As análises da  $R_0$  e  $R_i$  diárias pela aplicação do pacote MICE mostrou superestimava da  $R_0$  na série temporal do município de Arapiraca (-7,40), segundo o teste de PBIAS. Segundo Costa (2004), a superestimativa se deve aos erros de leitura de instrumento de medição ou mesmo na ausência de dados comparativos. Todavia, a análise série temporal da  $R_0$  e  $R_i$  diárias de acordo com os coeficientes de NSE e  $R^2$ , para Arapiraca (0,90 e 0,99).





**FIGURA 2.** Série temporal de radiação solar global ( $\text{KJ.m}^{-2}$ ) observado (em azul) e imputado (em preto) do município de Arapiraca no período de 2008 a 2016.

Na Figura 3 apresenta-se os valores dos índices estatísticos usados no diagrama 1:1. Os coeficientes R (Pearson), d (Willmott) e c (Camargo) foram superiores a 0,90 entre os dados  $R_0$  e  $R_i$ . O desempenho dos modelos de regressão linear simples (MRLS) apresentaram coeficientes r na categoria quase perfeita no município de Arapiraca ( $r = 0,99$ ). A análise dos parâmetros estatísticos EPE, RMSE e MAE indicou uma baixa dispersão do MRLS, em Arapiraca com EPE ( $113 \text{ KJ.m}^{-2}$ ), RMSE ( $137,49 \text{ KJ.m}^{-2}$ ) e MAE ( $111,47 \text{ KJ.m}^{-2}$ ).



**FIGURA 3.** Diagrama de espalhamento da radiação solar global ( $\text{KJ.m}^{-2}$ ) observado (em azul) e imputado (em preto) do município de Arapiraca no período de 2008 a 2016 com respectivos índices estatísticos (R, d, IC, EPE, RMSE, MAE, p-valor, Média O. e Média I.).

Para comprovar a existência de diferença entre as médias da radiação solar global observada e imputada ao nível de significância de 5%, foi aplicado o teste t de *Student* as séries temporais de dados de radiação solar global, em que ficou claro na Tabela 4, que a hipótese de nulidade para as seis regiões de estudo foram rejeitas para p. valor < 0,05, ou seja, existem diferenças não significativas (NS) nas médias observadas e imputadas de dados de radiação solar global.

**TABELA 4.** Teste de t de Student e comparações múltiplas das médias de Q<sub>g</sub>

Estação	Médias Observadas	Médias Imputadas	Teste de t de Student	p. valor do Teste t de Student
Arapiraca	1487,32 <sup>ab</sup>	1377,66 <sup>ab</sup>	12,36	1,15 x 10 <sup>-34</sup> (NS)

Legendas: Médias iguais - a, Médias diferentes – ab.

### CONCLUSÕES

A série de dados de Q<sub>g</sub> tem dados faltantes e, portanto o processo de imputação pelo método usado consegue preencher os dados ausentes de Q<sub>g</sub> com uma qualidade adequada, visto que há uma forte correlação (próximo de 1) entre os dados faltantes e os dados imputados.

Os parâmetros estatísticos (EPE, RMSE e MAE) temporais indicaram uma baixa dispersão do modelo de regressão linear simples em Arapiraca. O diagrama de dispersão da radiação solar observada e imputada são categorizados como ótimo e muito bom. O teste t de *Student* aplicado à série temporal de dados observados e imputados mostra que a hipótese de nulidade é rejeitada. Há diferenças não significativas (NS) nas médias observadas e imputadas, para p.valor < 0,05.

O uso do pacote MICE de imputação de dados via *software* ambiente R aplicado à série temporal da radiação solar global em Arapiraca é útil e, também devidamente aplicável a qualquer série temporal de dados climáticos.

### AGRADECIMENTOS

Os autores agradecem ao INMET por ceder gentilmente os dados de radiação solar global. O quarto autor agradece a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de Pós-Doutorado PNPd, junto ao Programa de Pós-Graduação em Tecnologia Ambiental (PGTA). O quinto autor agradecer a sua bolsa de Produtividade de Pesquisa do CNPq nível 2 sob processo 306410/2015-0.

### REFERÊNCIAS

BALTOKOSKI, V; TAVARES, M.H.F.; MACHADO, R.E; OLIVEIRA, M.P. Calibração de modelo para a simulação de vazão e de fósforo total nas sub-bacias dos rios Conrado e Pinheiro - Pato Branco (PR). **Revista Brasileira de Ciência do Solo**, v. 34, n. 1, p.253-261, 2010. Disponível em: <http://dx.doi.org/10.1590/S0100-06832010000100026>. doi: 10.1590/S0100-06832010000100026.

BRITO, T. T.; OLIVEIRA-JÚNIOR, J. F.; LYRA, G. B.; GOIS, G.; ZERI, M. Multivariate analysis applied to monthly rainfall over Rio de Janeiro state, Brazil. **Meteorology And Atmospheric Physics (Print)**, v. 129, n. 5, p. 469-478, 2016. Disponível em: <http://dx.doi.org/>

0.1007%2Fs00703-016-0481-x. doi: 10.1007%2Fs00703-016-0481-x.

BUUREN, S. V.; OUDSHOORN, K. G. MICE: Multivariate Imputation by Chained Equations in R. **Journal of Statistical Software**, v. 45, n. 3, 2011. Disponível em: <http://dx.doi.org/10.18637/jss.v045.i03>. doi: 10.18637/jss.v045.i03.

CAMARGO, A. P; SENTELHAS, P. C. Avaliação do desempenho de diferentes métodos de estimativas da evapotranspiração potencial no Estado de São Paulo, Brasil. **Revista Brasileira de Agrometeorologia**, v.5, n.1, p.89–97, 1997. Disponível em: <http://dx.doi.org/10.18637/jss.v045.i03>. doi: 10.18637/jss.v045.i03.

COSTA, E.; LEAL, P. A. M.; JUNIOR, R. R. C. Modelo de simulação da temperatura e umidade relativa do ar no Interior de estufa plástica. **Engenharia Agrícola**, Jaboticabal, v. 24, n.1, p. 57-67, 2004. Disponível em: <http://dx.doi.org/10.1590/S0100-69162004000100008>. doi: 10.1590/S0100-69162004000100008

GARCÍA-PEÑA, M.; ARCINIEGAS-ALARCÓN, S.; BARBIN, D. Imputação de Dados Climáticos Utilizando a Decomposição por Valores Singulares: Uma Comparação Empírica. **Revista Brasileira de Meteorologia**, v. 29, n. 4, p. 527-536, 2014. Disponível em: <http://dx.doi.org/10.1590/0102-778620130005>. doi: 10.1590/0102-778620130005

GOIS, G.; SOUZA, J. C. S.; DE OLIVEIRA-JÚNIOR, J. F.; DELGADO, R. C.; LYRA, G. B.; TEODORO, P. E. Models to estimate incident solar radiation on Seropédica, Rio de Janeiro. **Bioscience Journal (on line)**, v. 32, n. 2, p. 505-513, 2016. Disponível em: <https://doi.org/10.14393/BJ-v32n2a2016-29877>. doi: 10.14393/BJ-v32n2a2016-29877

GRÜND, S.; LÜDTKE, O.; ROBITZSCH, A. Multiple Imputation of Multilevel Missing Data: And Introduction to the R Package pan. **SAGE Open**. v. 6, 2016. Disponível em: <https://doi.org/10.1177/2158244016668220>. doi: 10.1177/2158244016668220

HARRELL, FE JR. **Regression modeling strategies with applications to linear models, logistic regression and survival analysis**. Springer–Verlag, New York, 2001.

HOPKINS, W. G. **Correlation coefficient: a new view of statistics**. 2009. Disponível em: <http://www.sportsci.org/resource/stats/correl.html>. Acesso em: 02 setembro de 2018.

IBGE – Instituto Brasileiro de Geografia e Estatística –, 2017. Mapa das Regiões Geográficas do Estado de Alagoas. Disponível em [ftp://geoftp.ibge.gov.br/organizacao\\_do\\_territorio/divisao\\_regional/divisao\\_regional\\_do\\_brasil/divisao\\_regional\\_do\\_brasil\\_em\\_regioes\\_geograficas\\_2017/mapas/33\\_regioes\\_geograficas\\_alagoas.pdf](ftp://geoftp.ibge.gov.br/organizacao_do_territorio/divisao_regional/divisao_regional_do_brasil/divisao_regional_do_brasil_em_regioes_geograficas_2017/mapas/33_regioes_geograficas_alagoas.pdf). Acesso em 25 de setembro de 2018.

KOWARIK, A.; TEMPL, M. Imputation with the R Package VIM. **Journal of Statistical Software**, v. 74, 2016. Disponível em: <https://doi.org/10.21256/zhaw-4233>. doi: 10.18637/jss.v074.i07

LAYANUN, V.; SUKSAMOSORN, S.; SONGSIRI, J. Missing-data imputation for solar irradiance forecasting in Thailand. 56<sup>th</sup> Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Kanazawa, p. 1234-1239, 2017.

LYRA, G. B.; OLIVEIRA-JÚNIOR, J. F.; GOIS, G.; CUNHA-ZERI, G.; ZERI, M. Rainfall variability over Alagoas under the influences of SST anomalies. **Meteorology and Atmospheric Physics (Print)**, v. 129, p. 157-171, 2017. Disponível em: <http://dx.doi.org/10.1007%2Fs00703-016-0461-1>. doi: 10.1007%2Fs00703-016-0461-1

LYRA, G. B.; ZANETTI, S. S.; SANTOS, A. A. R.; SOUZA, J. L.; LYRA, G. B.; OLIVEIRA-JÚNIOR, J. F.; LEMES, M. A. M. Estimation of monthly global solar irradiation using the Hargreaves-Samani model and an artificial neural network for the state of Alagoas in northeastern Brazil. **Theoretical and Applied Climatology**, v. 125, n. (3-4), p. 743-756, 2015. Disponível em: <https://link.springer.com/article/10.1007%2Fs00704-015-1541-8>. doi: 10.1007/s00704-015-1541-8

MORIASI, D. N.; ARNOLD, J. G.; LIEW, M. W. VAN; BINGER, R. L.; HARMEL, R. D.; VEITH, T. L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. **Transactions of the ASABE**, v.50, p.885-900, 2007.

NORMANDO, D.; TJÄDERHANE, L.; QUINTÃO, C. C. A. A escolha do teste estatístico – um tutorial em forma de apresentação em PowerPoint. **Dental Press Journal of Orthodontics**, v. 15, n. 1, p. 101-106, 2010. Disponível em: <http://dx.doi.org/10.1590/S2176-94512010000100012>. doi: 10.1590/S2176-94512010000100012.

NUNES, L.N.; KLÜCK, M.M.; FACHEL, J.M.G. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica. **Revista Brasileira de Epidemiologia**, v. 13, n. 4, p. 596-606, 2010. Disponível em: <http://dx.doi.org/10.1590/S1415-790X2010000400005>. doi: 10.1590/S1415-790X2010000400005.

OLIVEIRA JÚNIOR, J. F.; GOIS, G.; TERASSI, P. M. B.; SILVA JUNIOR, C. A.; BLANCO, C. J. C.; SOBRAL, B. S.; GASPARINI, K. A. C. Drought severity based on the SPI index and its relation to the ENSO and PDO climatic variability modes in the regions North and Northwest of the State of Rio de Janeiro - Brazil. **Atmospheric Research**, v. 212, n. 1, p. 91-105, 2018. Disponível em: <https://doi.org/10.1016/j.atmosres.2018.04.022>. doi: 10.1016/j.atmosres.2018.04.022.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org>, ISBN 3-900051-07-0, 2017.

SANTOS, J. C. **Análise da aplicação do modelo de domínio de conhecimento em administração territorial (LADM) ao cadastro territorial urbano brasileiro – estudo de caso para o município de Arapiraca - AL.** Programa de Pós-Graduação em Ciências Geodésicas e Tecnologias da Geoinformação, Universidade Federal de Pernambuco - UFPE, Recife, p.131, 2012.

SOBRAL, B. S.; OLIVEIRA-JÚNIOR, J. F.; GOIS, G.; PEREIRA JÚNIOR, E. R.; TERASSI, P. M. B. ; MUNIZ JÚNIOR, J. G. R.; LYRA, G. B. ; ZERI, M. Drought characterization for the state of Rio de Janeiro based on the annual SPI index: Trends, statistical tests and its relation with ENSO. *Atmospheric Research*, v. 220, n. 1, p. 141-154, 2019. Disponível em: <https://doi.org/10.1016/j.atmosres.2019.01.003>. doi: 10.1016/j.atmosres.2019.01.003

SOUZA, A. P.; SILVA, A. C.; TANAKA, A. A.; ULIANA, E. M., ALMEIDA, F. T.;KLAR, A. E.; GOMES, A. W. A. Global radiation by simplified models for the state of Mato Grosso, Brazil. **Pesquisa Agropecuária Brasileira**, v. 52, n.4, p.215-227, 2017. Disponível em: <http://dx.doi.org/10.1590/s0100-204x2017000400001>. doi: 10.1590/s0100-204x2017000400001

TEODORO, P. E.; DELGADO, R. C.; OLIVEIRA JÚNIOR, J. F.; GOIS, G.; TAYTSOHN, F. Incoming longwave radiation for the IPCC pessimistic scenario in the Legal Amazon. **Floresta e Ambiente**, v. 25, n.2, p. 1-8, 2018. Disponível em: <http://dx.doi.org/10.1590/2179-8087.009416> . doi: 10.1590/2179-8087.009416

TERASSI, P. M. B.; OLIVEIRA JÚNIOR, J. F.; GOIS, G.; GALVANI, E. Variabilidade do Índice de Precipitação Padronizada na Região Norte do Estado do Paraná Associada aos Eventos de El Niño-Oscilação Sul. **Revista Brasileira de Meteorologia**, v. 33, n. 1, p. 11-25, 2018. Disponível em: <http://dx.doi.org/10.1590/0102-7786331002> . doi: 10.1590/0102-7786331002

VIEIRA, S. A.; OSORIO, D. M. M.; QUEVEDO, D. M., ADAM, K. N.; PEREIRA, M. A. F. Metodologia de Imputação de Dados Hidrometeorológicos para Análise de Séries Históricas – Bacia do Rio dos Sinos, RS, Brasil. **Revista Brasileira de Climatologia**, v. 23, n. 1, p. 189-204, 2018. Disponível em: <http://dx.doi.org/10.5380/abclima.v23i0.56219>. doi: 10.5380/abclima.v23i0.56219

VIOLA, M. R.; MELLO, C.R.; ACERBI JUNIOR, F.W.; SILVA, A.M. Modelagem hidrológica na bacia hidrográfica do Rio Aiuruoca, MG. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n.5, p. 581–590, 2009. Disponível em: <http://dx.doi.org/10.1590/S1415-43662009000500011>. doi: 10.1590/S1415-43662009000500011

WILLMOTT, C. J. **On the validation of models. Physical Geography**, v.2, n. 2, p. 184-194, 1981.

ZHANG, Z. Missing data imputation: focusing on single imputation. **Ann Transl Med**. V. 4, n. 1, 2016. Disponível em: <http://dx.doi.org/10.3978/j.issn.2305-5839.2015.12.38> doi: 10.3978/j.issn.2305-5839.2015.12.38