APLICAÇÃO E AVALIAÇÃO DE TÉCNICAS PARA O PREENCHIMENTO DE FALHAS DE DADOS PLUVIOMÉTRICOS EM ANOS HABITUAIS, SECOS E CHUVOSOS

APPLICATION AND EVALUATION OF TECHNIQUES TO FILL IN THE GAPS OF RAINFALL DATA IN USUAL, DRY AND RAINY YEARS

APLICACIÓN Y EVALUACIÓN DE TÉCNICAS PARA EL LLENADO DE FALLOS DE DATOS LLUVIOSOS EN AÑOS HABITUALES, SECOS Y LLUVIOSOS

Thiago Alves de Oliveira
Doutorando em Geografia Física pela Universidade de São Paulo (USP).
thiago.a.oliveira@usp.br

Fabio Sanches
Doutor em Geografia. Universidade Federal de Juiz de Fora (UFJF).
fabio.sanches@ufjf.br

Cássia de Castro Martins Ferreira
Doutora em Geografia. Universidade Federal de Juiz de Fora (UFJF).
cassia.castro@ufjf.edu.br

RESUMO

Séries de dados climáticos geralmente apresentam eventuais falhas em seus conjuntos. Este trabalho objetivou preencher falhas de dados pluviométricos em Juiz de Fora avaliando sua consistência e qual técnica melhor representa o comportamento das chuvas mensais. Dados das estações de Lima Duarte, Chapéu D'Uvas e Mar de Espanha foram utilizados no preenchimento das falhas de Juiz de Fora. No preenchimento foram utilizadas técnicas de Regressão Linear Simples e de Ponderação Regional. Após, os conjuntos de dados foram avaliados pelos Coeficientes de Determinação (R^2) e pela Correlação de Pearson (r). Testes Dupla-Massa foram aplicados para avaliar possíveis erros sistemáticos. A etapa seguinte avaliou a capacidade das técnicas em sintetizar dados observados e para isso, foram selecionados anos Habituais, Secos e Chuvosos. Os dados sintetizados foram comparados aos observados e avaliados em sua homogeneidade pelo Teste *t* de *Student*, e sua dispersão pelo Erro Médio Absoluto, Erro Relativo Médio e da Raiz do Erro Quadrado Médio. Os resultados demonstraram elevada qualidade no preenchimento por todas as técnicas utilizadas, destacando a Ponderação Regional por ter apresentado os menores erros sistemáticos para os anos Habituais, Secos e Chuvosos.

Palavras–chave: Chuvas; Testes estatísticos; Juiz de Fora; Zona da Mata Mineira; Anos-padrão.

ABSTRACT

Climatic data series usually present eventual gaps in their sets. This work aimed to fill gaps in rainfall data in *Juiz de Fora*, evaluating its consistency and which technique best represents the behavior of monthly rainfall. Data from the *Lima Duarte*, *Chapéu D'Uvas*, and *Mar de Espanha* stations were used to fill in the gaps in *Juiz de Fora*. Simple Linear Regression and Regional Weighting techniques were used in filling in the gaps. Then the data sets were evaluated by the Coefficients of Determination (R^2) and Pearson's Correlation (r). Double-Mass tests were applied to evaluate possible systematic errors. The next step evaluated the ability of the techniques to synthesize observed data and for this, Habitual, Dry and Rainy years were selected. The synthesized data were compared to the observed data and evaluated for homogeneity using Student's t-test, and dispersion using the Absolute Mean Error, Relative Mean Error, and Root Mean Square Error. The results showed high quality in the filling of the gaps by all the techniques used, with Regional Weighting standing out as having the lowest systematic errors for the Usual, Dry and Rainy years.

Keywords: Rainfall; Statistical tests; Juiz de Fora; Zona da Mata Mineira; Standard years.

RESUMEN

Las series de datos climáticos suelen presentar eventuales lagunas en sus conjuntos. Este trabajo tuvo como objetivo llenar las lagunas en los datos de precipitación en *Juiz de Fora*, evaluando su consistencia y qué técnica representa mejor el comportamiento de la precipitación mensual. Los datos de las estaciones de *Lima Duarte*, *Chapéu D'Uvas* y *Mar de Espanha* se utilizaron para rellenar las lagunas de *Juiz de Fora*. Se utilizaron las técnicas de regresión lineal simple y de ponderación regional. Posteriormente, los conjuntos de datos se evaluaron mediante los coeficientes de determinación (R^2) y la correlación de Pearson (r). Se aplicaron pruebas de doble masa para evaluar los posibles errores sistemáticos. En el siguiente paso se evaluó la capacidad de las técnicas para sintetizar los datos observados y para ello se seleccionaron los años Habitual, Seco y Lluvioso. Los datos sintetizados se compararon con los datos observados y su homogeneidad se evaluó mediante la prueba t de Student, y su dispersión mediante el Error Medio Absoluto, el Error Medio Relativo y el Error Medio Cuadrático. Los resultados mostraron una alta calidad en el relleno de las lagunas por todas las técnicas utilizadas, destacando la Ponderación Regional por haber presentado los menores errores sistemáticos para los años Usual, Seco y Lluvioso.

Palabras clave: Precipitación; Pruebas estadísticas; Juiz de Fora; Zona da Mata Mineira; Años estándar.

INTRODUÇÃO

Nas últimas décadas, sobretudo a partir dos anos 1980, frente a ocorrência de fenômenos atmosféricos excepcionais e de desastres socioambientais associados à dinâmica atmosférica, os estudos em climatologia assumiram destaque especial tanto em eventos de comunidades específicas na área (meteorologia, geografia, hidrologia, agronomia, engenharia ambiental etc.), quanto às demandas por organismos públicos no que se referem aos problemas em questão.

Para qualquer estudo desenvolvido em climatologia, para além da adoção de técnicas específicas na aplicação em determinados estudos sobre clima, a existência e disponibilidade de dados meteorológicos pretéritos, com a devida qualidade e consistência e que permitam obter longas séries de dados temporais, mostra-se cada vez mais necessários para atender demandas nesses estudos.

O conhecimento sobre a variabilidade de determinado fenômeno climático (estiagens, de eventos pluviométricos intensos, estimativas de retornos etc.), até mesmo possíveis modificações nos aspectos climáticos passam, essencialmente, pela minuciosa análise dessas longas séries de dados temporais. Entretanto, um problema recorrente enfrentado pela maioria dos pesquisadores do clima, refere-se em possuir tais conjuntos de dados.

Infelizmente, parte das informações meteorológico-climáticas disponíveis nos principais bancos de dados brasileiros, sobretudo no Hidroweb[1] da Agência Nacional de Águas (ANA) e no Banco de Dados Meteorológicos para Ensino e Pesquisa[2] do Instituto Nacional de Meteorologia (BDMEP/INMET), apresenta falhas ou interrupções em suas séries ou conjuntos de dados. Essas falhas, muitas vezes, se devem a danos aos equipamentos, período de manutenção, ausência de observador (férias, aposentadoria, por exemplo), falhas na transmissão de dados etc.

Frente a esse problema, os pesquisadores do clima utilizam-se de uma gama de técnicas as quais, quando aplicáveis, buscam preencher tais lacunas junto aos dados

[1] http://www.snirh.gov.br/hidroweb/apresentacao
[2] http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep

originais. Porém, o uso dessas técnicas (ou de apenas uma), depende diretamente da disponibilidade dos dados em estações próximas e das características ambientais de seu entorno.

A proximidade (ou não) de outros postos/estações meteorológicas, suas condições topográficas (altimetria, posição das vertentes etc.), condições ambientais no entorno desses postos/estações (em áreas urbanas, áreas rurais, cobertura vegetal, parques etc.), existência de dados em períodos concomitantes, entre outros, são parâmetros que merecem profunda consideração quando se pretende adotar procedimentos para o preenchimento de falhas existentes, como destacam Amador e Silva (2020).

Dessa forma, a adoção da técnica mais adequada ou de um conjunto de técnicas depende, diretamente, das condições preliminares dos conjuntos de dados a serem utilizados.

A literatura recente demonstra que há um amplo conjunto de técnicas consagradas como o uso das Médias (MILANESI e GALVANI, 2006; VENTURA, et al. 2016; JUNQUEIRA et al. 2018), das Médias Móveis (VENTURA, et al. 2016; CABRAL JÚNIOR e LUCENA, 2019; CABRAL JUNIOR et al. 2019), da Regressão Linear (NASCIMENTO, et al. 2009; OLIVEIRA, et al. 2010; SANCHES et al. 2014a; 2014b; PECHUTTI e SANT'ANNA NETO, 2016; VENTURA, et al. 2016; JUNQUEIRA et al. 2018; RUEZZENE et al. 2020), da Regressão Linear Múltipla (VENTURA, et al. 2016; MELLO et al. 2017; JUNQUEIRA et al. 2018), da Ponderação Regional (MILANESI e GALVANI, 2006; OLIVEIRA, et al. 2010; PECHUTTI e SANT'ANNA NETO, 2016; MELLO et al. 2017; DIAZ, et al. 2018; JUNQUEIRA, et al. 2018; AMADOR e SILVA, 2020; RUEZZENE et al. 2020), da Ponderação Regional com base em Regressão Linear (OLIVEIRA, et al. 2010; MELLO et al. 2017; JUNQUEIRA et al. 2018), do Inverso Ponderado da Distância (IDW) (JUNQUEIRA et al. 2018; RUEZZENE et al. 2020) entre outras.

Em tese, após o preenchimento das falhas é necessário adotar procedimentos de avaliação quanto à qualidade de tais preenchimentos. O Teste Dupla-Massa (DM) é uma técnica amplamente recomendada para se verificar a consistência do (s) preenchimento (s) em séries temporais (BERTONI e TUCCI, 2007). Sanches et al.

(2014a; 2014b), Junqueira et al. (2018) entre outros fizeram uso do DM em seus trabalhos com resultados consistentes. Para Bertoni e Tucci (2007), a versatilidade do Teste Dupla-Massa permite que seja utilizado na avaliação e séries de dados anual e mensal.

Da mesma forma, a verificação de possíveis erros sistemáticos em função do preenchimento de tais falhas pode ser feita por meio da análise tanto do conjunto central, quanto da dispersão de tais valores. O Coeficiente de Determinação e o Coeficiente de Correlação de Pearson foram utilizados por Nascimento et al. (2009), Oliveira et al. (2010), Sanches et al. (2014a; 2014b), Pechutti e Sant'Anna Neto (2016), Ventura et al. (2016), Diaz et al. (2018) e Junqueira et al. (2018) para a avaliação do conjunto central entre os dados preenchidos e os dados observados. Araújo et al. (2017) e Diaz et al. (2018) recomendam ainda que, além desses coeficientes (de determinação e de correlação) seja adotado o Teste *t* de *Student* para avaliar a homogeneidade de ambos os conjuntos.

Por sua vez, testes como o Erro Médio Absoluto (EMA), o Erro Relativo Médio (ERM) e a Raiz do Erro Quadrático Médio (REQM) avaliam o grau de dispersão do conjunto de dados após o preenchimento. Tais técnicas foram amplamente utilizadas por Ventura et al. (2016), Diaz et al. (2018), Junqueira et al. (2018) e Quina et al. (2020) em suas pesquisas com resultados igualmente consistentes e robustos.

Dessa forma, o presente trabalho tem como objetivo a adoção e comparação de técnicas para o preenchimento de falhas em dados pluviométricos para Juiz de Fora (MG), bem como a avaliação da qualidade dos dados preenchidos considerando anos classificados como Habituais, Secos e Chuvosos.

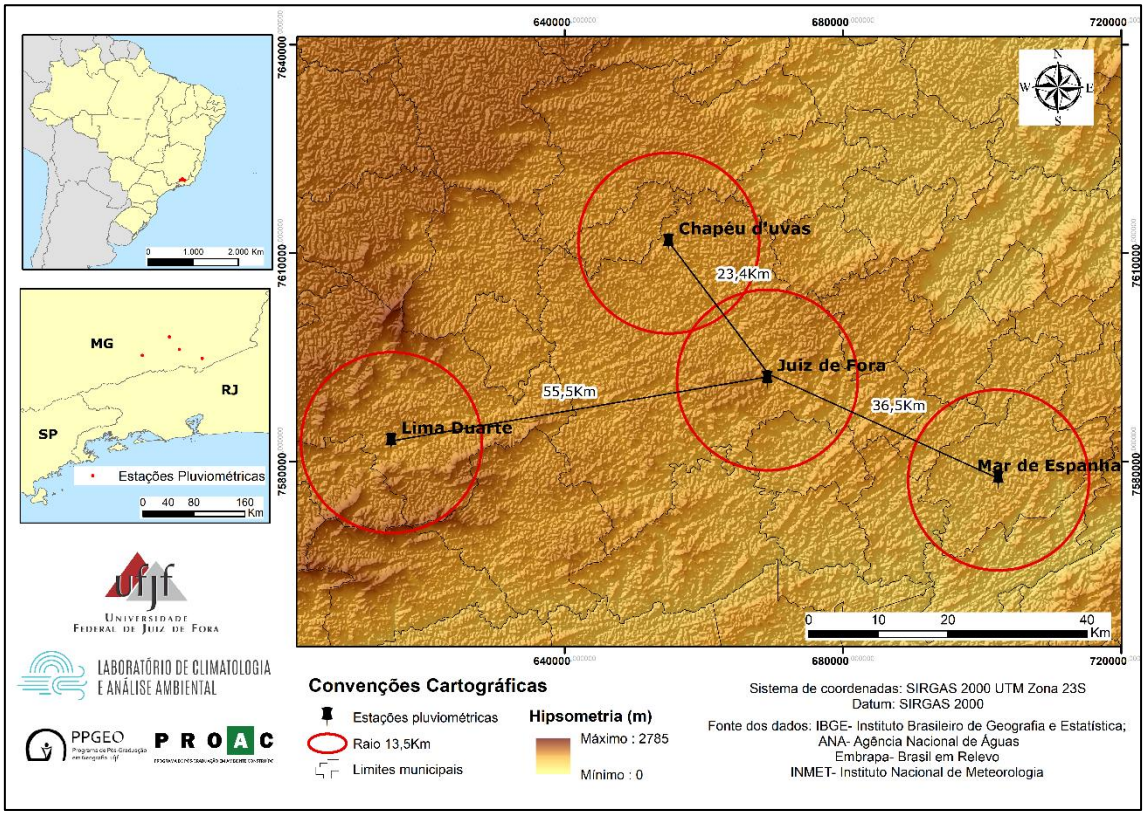
MATERIAL E MÉTODO

Localização da área de estudo

Juiz de Fora está localizada na região Zona da Mata do estado de Minas Gerais (Figura 1), a qual foi caracterizada em Ab'Sáber (2003) por morrarias arredondadas recobertas por vegetação de Mata Atlântica, pertencentes ao Domínio dos Mares de Morros. As altitudes da região variam entre 650 e 950 m, peculiares à

localização das demais estações meteorológicas utilizadas (Lima Duarte, Chapéu D'Uvas e Mar de Espanha) (EDUARDO, 2018).

Figura 1 - Localização das Estações Meteorológicas de Lima Duarte (LD), Chapéu D'Uvas (CH) e Mar de Espanha (ME) e suas respectivas distâncias lineares em relação a Estação Meteorológica de Juiz de Fora.



Para Ferreira (2012) e Oliveira et al (2020), o regime pluviométrico na região de estudo (parte da Zona da Mata Mineira) apresenta duas estações distintas: uma chuvosa (out/mar) e outra seca (abri/set), com quantitativos anuais em torno de 1500 mm, típicas de uma condição de tropicalidade.

Dados utilizados

No presente trabalho foram utilizados dados pluviométricos mensais das Estações Meteorológicas do INMET de Juiz de Fora (INMET: 83692 e A518), no período de 1945-2020 e das estações meteorológicas da ANA de Lima Duarte (ANA: 2143019),

no período de 1944-2020, de Chapéu D'Uvas (ANA: 2143020), no período de 1950-2020 e de Mar de Espanha (ANA: 2143013), no período de 1943-2020.

Ao se analisar os totais pluviométricos mensais da Estação de Juiz de Fora observou-se falhas nos meses de janeiro a dezembro nos anos de 1956, 1957, 1958, 1959, 1960, 1961 e 1962.

Já nos anos de 1964 e 1965, as falhas foram observadas apenas nos meses de março, bem como no ano de 1967 (mês de abril), 1971 (mês de fevereiro), 1972 (de janeiro à julho), 1973 (de março à julho), 1991 (de março à abril), 1992 (de março à julho) e 1997 (de agosto à dezembro), correspondendo ao todo, cerca de 13,3% do conjunto total de dados para Juiz de Fora.

Técnicas de preenchimento de dados

Dadas as características dos conjuntos de dados pluviométricos disponíveis (estações meteorológicas próximas, quantidade de estações, série temporal disponível), foram adotadas duas técnicas para o preenchimento das falhas existentes: a Regressão Linear simples e a Ponderação Regional.

Regressão Linear simples:

$$y = \alpha x + \beta \tag{1}$$

Onde y é o valor da falha a ser preenchida (variável dependente), *x* é o valor do posto/estação vizinha, α é o coeficiente angular da reta de regressão e β o coeficiente linear da reta.

Ponderação Regional:

$$y = \frac{1}{n}\left[\frac{x_1}{\bar{x}_1} + \frac{x_2}{\bar{x}_2} + \frac{x_3}{\bar{x}_3} \ldots \frac{x_i}{\bar{x}_i}\right] . \bar{y} \tag{2}$$

Onde *y* é a variável que se pretende encontrar (falha a ser preenchida), *n* é o número de postos/estações utilizadas para o processo de preenchimento, *x* são os valores

dos postos/estações vizinhas, $\bar{x}$ é a média de precipitação de cada posto e $\bar{y}$ a média de precipitação do posto/estação que possuía falha a ser preenchida.

Técnicas de avaliação de consistência

Após a etapa do preenchimento das falhas, deve-se analisar a consistência do conjunto de dados preenchidos. Neste trabalho, a primeira técnica utilizada para a avaliação dessa consistência foi a aplicação do Teste Dupla-Massa (DM) *cf.* Bertoni e Tucci (2007), Sanches et al. (2014a; 2014b) e Junqueira et al. (2018).

No DM, os dados da série histórica preenchida devem ser proporcionais à série à média dos postos/estações vizinhas (variável de controle), devendo os pontos alinhar-se segundo uma reta. Para isso, plota-se em um gráfico cartesiano os totais acumulados de precipitação (variável de controle) compondo o conjunto de dados do eixo das abscissas (*eixo x*) e o conjunto dos totais mensais acumulados a qual se pretende verificar (variável de interesse) no eixo das ordenadas (*eixo y*) (BERTONI e TUCCI, 2007).

A seguir, a verificação semelhança (homogeneidade) entre os dados preenchidos e os observados se deu por meio da aplicação de um conjunto de técnicas estatísticas para sua validação. No primeiro passo o resultado dos dados preenchidos foram comparados com os dados observados por meio da aplicação dos Coeficientes de Determinação (R^2) e de Correlação de Pearson (r).

O Coeficiente de Determinação (Equação 3) consiste na razão entre a soma de quadrados da regressão e a soma de quadrados total, variando entre 0 (zero) e 1 (um). Quanto mais próximo a 1, mais explicativo é o modelo e melhor é o ajuste dos dados da amostra (SPIEGEL, 1993).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \right)^2 \quad (3)$$

O Coeficiente da Correlação de Pearson (Equação 4) é usado para analisar o desempenho das técnicas de preenchimento de falhas, caracterizando-se por Muito Forte (r entre 0,90 e 1,0), Forte (r entre 0,70 a 0,89), Moderado (r entre 0,50 até 0,69) e Fraco (r entre 0,30 e 0,49) e Desprezível (r entre 0 e 0,29) (HINKLE et al. 2003)

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (4)$$

Tanto na Equação 3 como na Equação 4, *n* consiste no número de observações, x_i e y_i correspondem aos valores de ambos os conjuntos de dados (observados e preenchidos) e, $\bar{x}$ e $\bar{y}$ suas respectivas médias.

A partir da caracterização dos anos Habituais, Secos e Chuvosos propostos por Oliveira et al., (2020), foram selecionados os anos de 1964, 1967, 1999 e 2007 como anos Habituais, 1953, 1970 e 2014 como anos Secos e 1983, 1994 e 2008 como anos Chuvosos, totalizando dez anos para simulação dos procedimentos técnicos (aplicação dos testes) para preenchimento de falhas.

Assim como no trabalho de Pechutti e Sant'Anna Neto (2016), tal procedimento foi adotado como forma de verificar da capacidade das técnicas aplicadas em representar as falhas preenchidas nos dados de Juiz de Fora

Os resultados obtidos por meio das regressões lineares e da ponderação regional foram avaliados por meio do Teste *t* de *Student* (entre o simulado e o observado) e dos testes de avaliação dos valores dispersivos (Erro Médio Absoluto - EMA, o Erro Relativo Médio – ERM e a Raiz do Erro Quadrático Médio - REQM).

A aplicação do Teste *t* de *Student* para verificar a homogeneidade de ambos os conjuntos de dados (observado e simulado) após a fase de preenchimento *cf.* recomendam Araújo et al. (2017) e Diaz et al. (2018).

Por tratar-se de um teste de hipótese, inicialmente assume-se que ambos os conjuntos de dados são estatisticamente homogêneos (H_0), contra uma hipótese alternativa (H_1), na qual os conjuntos de dados mostram-se heterogêneos (SPIEGEL, 1993).

Sendo assim, tem-se que:

$$t = \frac{\mu_1 - \mu_2}{\sigma \sqrt{1/N_1 + 1/N_2}} \quad (5)$$

onde

$$\sigma = \sqrt{\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}} \quad (6)$$

e

$$s^2 = \frac{(\sum_{i=1}^{n} X_i - \mu)^2}{N - 1} \quad (7)$$

Dessa forma, considerando-se que μ_1 e μ_2 correspondem às médias mensais dos dados observados e simulados, que N_1 e N_2 são as amostras de ambos os conjuntos de dados, σ é o desvio padrão entre os conjuntos de dados e s_1^2 e s_2^2 as respectivas variâncias de cada série.

Para o grau de liberdade da distribuição *t*, adotou-se que:

$$V = N_1 + N_2 - 2 \quad (8)$$

A partir do resultado da estatística *t*, em um nível de confiança de 95% (*p-value*: 0,05), compete aceitar (ou refutar) a hipótese H_0 caso seu escore (z) esteja entre -2,26 e 2,26 ($-2{,}26 \leq z \leq 2{,}26$).

Avaliadas as consistências do conjunto central dos dados, passou-se para a avaliação da dispersão do conjunto de dados simulados. Para isso, foram aplicadas as técnicas do Erro Médio Absoluto (EMA), o Erro Relativo Médio (ERM) e a Raiz do Erro Quadrático Médio (REQM) de acordo com Ventura et al. (2016), Junqueira et al. (2018) e Quina et al. (2020).

O EMA (Equação 9) consiste na diferença entre os dados observados e os dados simulados, indicando que, quanto menor for o seu valor (valor de EMA), mais próximo os valores simulados são dos valores observados.

$$EMA = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \quad (9)$$

Na Equação 9, *n* consiste no número de amostras, x_i o valor do estimador (falhas simuladas) e $\overline{\boldsymbol{x}}$ a média dos valores de todas as amostras (dados observados).

O ERM (Equação 10) indica a magnitude média do erro estimado com os melhores valores, aqueles mais próximos de zero (0). Assim, na Equação 10, *n* corresponde ao número de amostras, x_i é o valor fornecido pela falha simulada/preenchida para a *i*-ésima amostra, $\overline{\boldsymbol{x}}$ é a média dos valores de todas as amostras preenchidas, e $\dot{\boldsymbol{X}}\boldsymbol{i}$ é o valor correto que deve ser fornecido pelo dado observado, para a amostra em questão.

$$ERM = \frac{\sum_{i=1}^{n}|x_i - \dot{x}_i|}{\sum_{i=1}^{n}|x_i - \bar{x}|} \quad (10)$$

Por fim, o REQM (Equação 11) avalia o grau de espalhamento dos dados simulados comparados aos dados observados, onde *n* corresponde ao número de amostras, x_i é o valor fornecido pela técnica de preenchimento de falha e $\overline{\boldsymbol{x}}$ a média dos valores de todas as amostras dos valores observados.

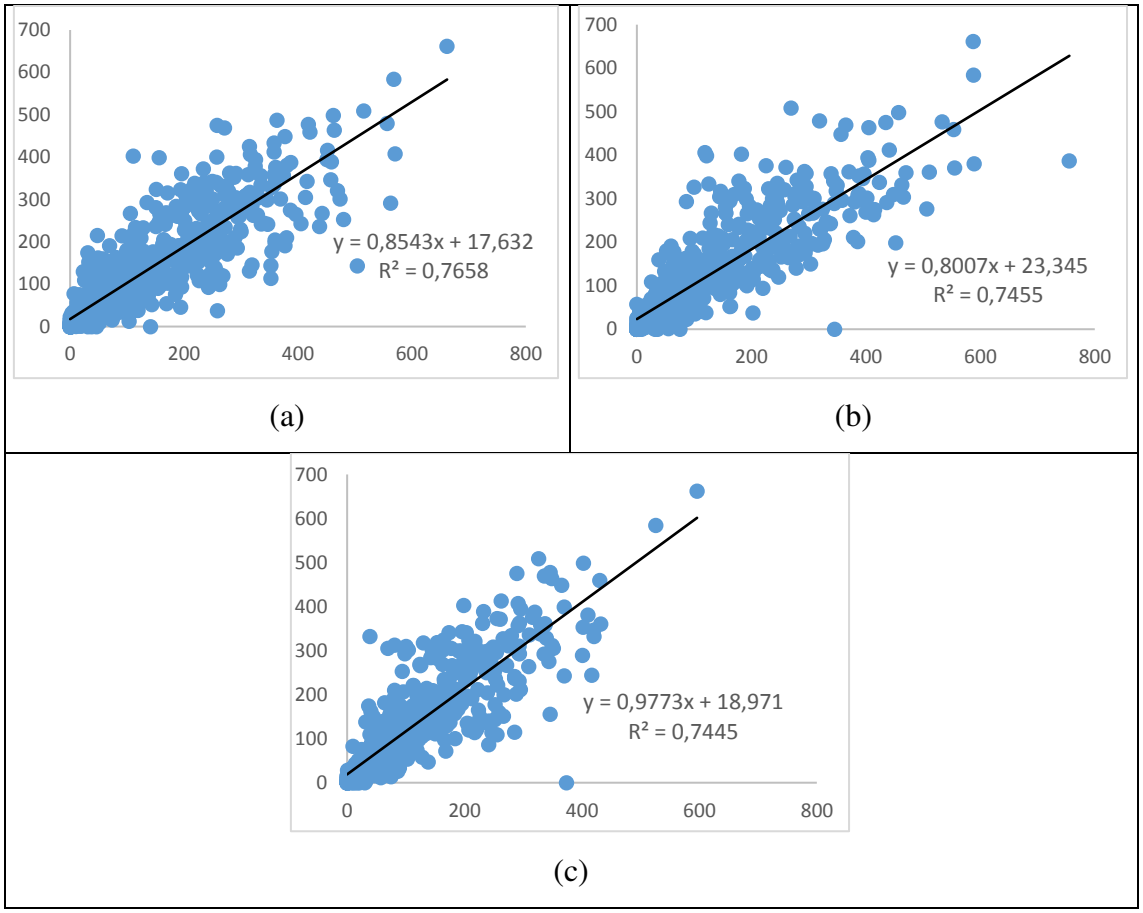
$$REQM = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (11)$$

RESULTADOS E DISCUSSÃO

Inicialmente os procedimentos consistiram em correlacionar os dados das estações meteorológicas de Lima Duarte, Chapéu D'Uvas e Mar de Espanha com os dados da estação meteorológica de Juiz de Fora.

A Figura 2 traz os gráficos de correlação de dados de pluviosidade (mm) de cada estação utilizada com os dados da estação de Juiz de Fora. Observa-se, também, suas equações de regressão (utilizadas para estimar as falhas nos dados de Juiz de Fora), bem como seus coeficientes de determinação (R^2).

Figura 2 - Gráficos de correlação entre os dados de pluviosidade (mm) observados e preenchidos para (a) Lima Duarte e Juiz de Fora, (b) Chapéu D'Uvas e Juiz de Fora e (c) Mar de Espanha e Juiz de Fora.



O valor do R^2 é utilizado para explicar, percentualmente, o grau de dispersão dos dados correlacionados em função de uma reta (reta de regressão). Quanto menor a dispersão dos dados, maior é esse percentual.

Dessa forma, a partir dos valores obtidos para tal parâmetro (Tabela 1), verifica-se que embora os R^2 das três estações utilizadas sejam elevados (superiores a

70% de explicação), o valor do coeficiente obtido entre as estações e Lima Duarte e Juiz de Fora foi o maior, o qual explica cerca de 77% dos dados.

Tabela 1 - Resultado dos coeficientes de determinação (R^2) e da Correlação de Pearson (r) a partir das Regressões Lineares aplicadas.

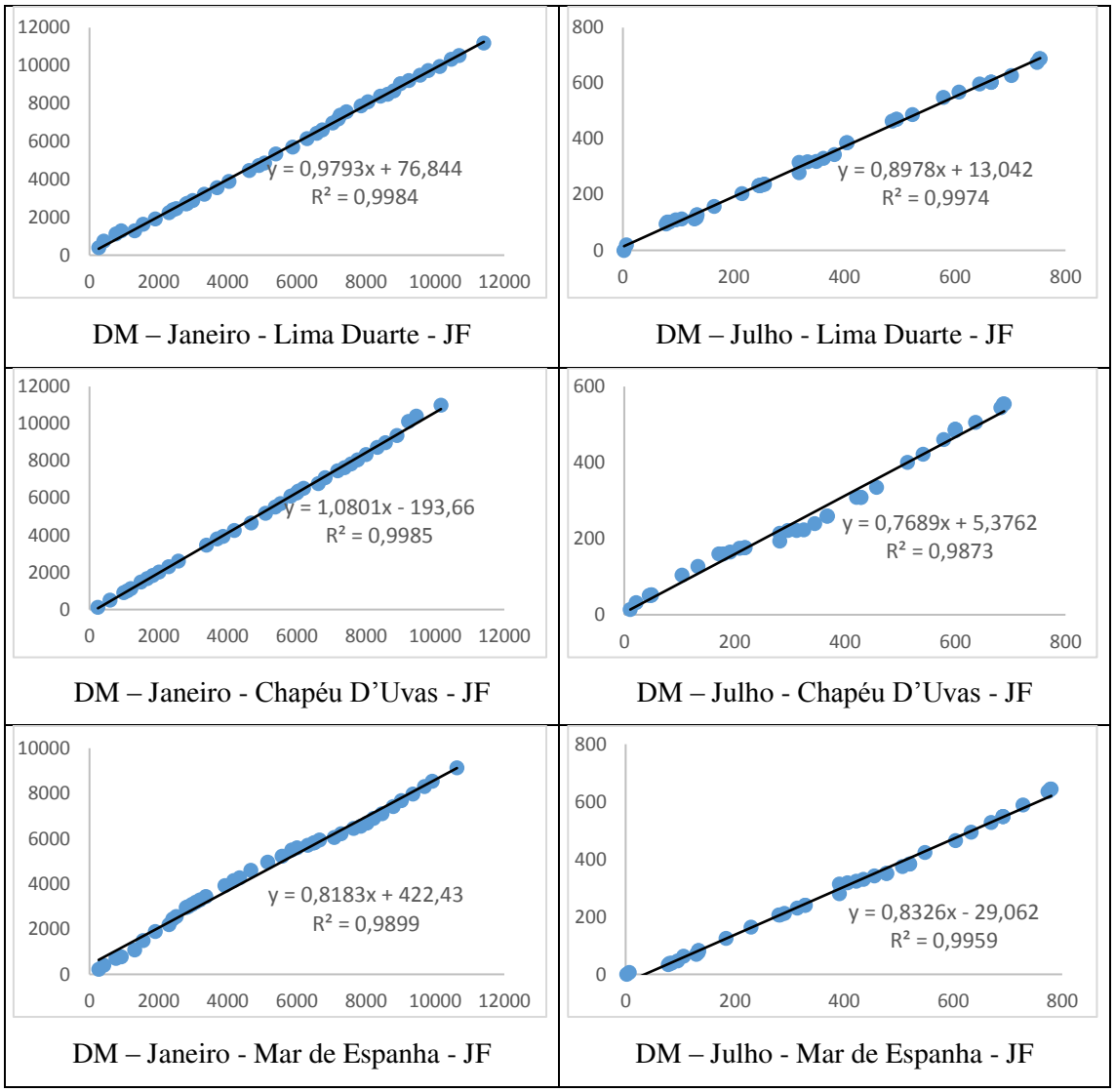
Estações correlacionados	R^2	r
Lima Duarte (LD) e Juiz de Fora (JF)	0,77	0,88
Chapéu D'Uvas (CH) e Juiz de Fora (JF)	0,75	0,86
Mar de Espanha (ME) e Juiz de Fora (JF)	0,74	0,86

Igualmente, os resultados da Correlação de Pearson (r) entre as estações de Lima Duarte, Chapéu D'Uvas e Mar de Espanha com os dados da estação de Juiz de Fora (Tabela 1) podem ser consideradas como "correlações fortes" (entre 0,70 e 0,89) de acordo com Hinkle et al (2003). No entanto, novamente, o Coeficiente de Pearson obtido entre os dados de Lima Duarte e Juiz de Fora (r=0,88) mostrou-se superior aos coeficientes das demais estações correlacionadas.

Após o preenchimento das falhas por meio das Regressões Lineares, os resultados do Teste Dupla-Massa em escala mensal indicaram valores de R^2 entre 0,97 e 0,99, demonstrando que, em todos os meses e em todas as técnicas aplicadas, não houve erros sistemáticos nos preenchimentos analisados.

A Figura 3 mostra, como exemplos desses resultados, o DM dos meses de janeiro e julho para as três estações.

Figura 3 - Resultado do Teste Dupla-Massa (DM) para os meses de Janeiro e Julho.



O passo seguinte consistiu na verificação da capacidade dos modelos de Regressão Linear e de Ponderação Regional em representar a pluviosidade das falhas preenchidas, buscando preencher falhas simuladas em anos considerados como Habituais, Secos e Chuvosos.

A classificação dos anos padrão (Habituais, Secos e Chuvosos) foi obtida à partir da metodologia descrita em Galvani e Luchiari (2012), selecionando-se, assim, os anos de 1964, 1967, 1999 e 2007 como sendo Habituais; 1953, 1970 e 2014 como Secos e, 1983, 1994 e 2008 como Chuvosos.

O resultado da avaliação da homogeneidade (Teste *t* de *Student*) entre os dados observados e simulados estão descritos na Tabela 2.

Tabela 2 - Resultados dos escores "z" da estatística de *Student* entre os dados observados e simulados pelos modelos de Regressão Linear e pela Ponderação Regional.

	Jan	**Fev**	**Mar**	**Abr**	**Mai**	**Jun**	**Jul**	**Ago**	**Set**	**Out**	**Nov**	**Dez**
Reg. LD/JF	-0,53	-2,10	0,73	0,16	1,83	-0,13	-0,82	-0,98	-0,13	0,46	-0,50	-0,30
Reg. CH/JF	0,15	-0,99	1,48	-0,37	0,30	-0,54	0,69	-1,52	-1,66	0,38	-0,26	1,80
Reg. ME/JF	-1,36	1,06	0,79	-0,17	-0,58	-0,85	-0,6	-0,05	0,29	1,49	-0,68	-0,83
Pond. Reg.	-0,55	-0,07	1,47	-0,05	1,24	0,14	1,41	-0,50	0,15	0,42	-1,14	0,23

Lima Duarte (LD); Chapéu D'Uvas (CH); Mar de Espanha (ME); Juiz de Fora (JF).

Nível de Confiança de 95%.

Os resultados da estatística de *Student* (Tabela 2) demonstram que em todas as simulações realizadas (Regressões Lineares e Ponderação Regional), os conjuntos de dados foram considerados estatisticamente homogêneos, aceitando-se a hipótese H_0 ($-2{,}26 \leq z \leq 2{,}26$), em um nível de confiança de 95%.

Dessa forma, é possível considerar que as técnicas de preenchimento de falhas adotadas e aplicadas representaram, com elevado grau de confiança, o comportamento das precipitações nos anos considerados Habituais, Secos e Chuvosos. Sendo assim, cabe avaliar o nível dos erros sistemáticos (valores dispersivos) para se identificar o modelo de preenchimento de falhas mais eficiente para este conjunto de dados.

A Tabela 3 traz os parâmetros da avaliação dos erros sistemáticos aplicados a ambos os conjuntos de dados (observados e simulados), os quais indicam resultados altamente satisfatórios com a capacidade de preenchimento das falhas.

Tabela 3 - Resultados dos coeficientes de erro dos valores dispersivos a partir dos preenchimentos pelas Regressões Lineares e pela Ponderação Regional.

Tipo de preenchimento adotado	**Coeficientes de Erro**		
	EMA	**ERM**	**REQM**
Regressão Linear - Lima Duarte (LD)/Juiz de Fora	0,61	0,83	12,9
Regressão Linear – Chapéu D'Uvas (CH)/Juiz de Fora	-0,70	0,74	13,2
Regressão Linear – Mar de Espanha (ME)/Juiz de Fora	0,03	-0,32	10,8
Ponderação Regional (LD/CH/ME)	-0,10	-0,02	9,2

Sob a perspectiva estatística, a média dos erros individuais é o primeiro passo na identificação das diferenças entre os conjuntos de dados (observados e simulados) e que tais erros (individuais) positivos e negativos se cancelam no somatório (QUINA et al, 2020).

Dessa forma, o EMA contorna esse problema uma vez que é menos afetado por valores anomalamente extremos (*outliers*), assumindo que resultados próximos a zero (0) são considerados como condição perfeita.

Este mesmo princípio deve ser adotado ao se avaliar os resultados do ERM e da REQM, uma vez que, à medida que seus valores aumentam, as diferenças entre os valores observados e simulados (falhas preenchidas) igualmente aumentam.

Sendo assim, os resultados dos conjuntos dispersivos entre os valores observados e os valores simulados para os anos Habituais, Secos e Chuvosos (Tabela 3), sugerem que, neste caso, a técnica da Ponderação Regional mostrou-se mais eficiente, pois apresentou os menores erros sistemáticos.

CONSIDERAÇÕES FINAIS

A disponibilidade e a qualidade dos dados de pluviosidade das estações meteorológicas relativamente próximas à estação meteorológica de Juiz de Fora possibilitou a adoção de dois conjuntos de técnicas estatísticas (Regressão Linear simples e a Ponderação Regional) para o preenchimento de falhas existentes na série de Juiz de Fora. Tais condições (três estações próximas, com séries longas e com dados completos) são relativamente raras de se encontrar nos repositórios de dados de pluviosidade.

As Regressões Lineares simples são comumente aplicadas como ferramenta para o preenchimento de falhas em conjuntos de dados pluviométricos. A Ponderação Regional é aplicada em casos em que há um conjunto de dados concomitantes e disponíveis em pelo menos duas estações meteorológicas próximas.

Neste caso, foi possível preencher as falhas avaliar a qualidade tanto das Regressões Lineares quando da Ponderação Regional por meio de um amplo conjunto de

ferramentas estatísticas de avaliação (Teste Dupla-Massa, Coeficientes de Determinação, Correlação de Pearson, Teste *t* de *Student*, EMA, ERM e REQM).

Embora todas as técnicas adotadas aos dados de todos as estações tenham demonstrado elevada qualidade e significância no preenchimento das falhas na série de Juiz de Fora, a técnica da Ponderação Regional mostrou-se mais eficiente por ter apresentado o menor conjunto de erros sistemáticos aos dados preenchidos.

Cabe reforçar, aqui, a ideia de que a disponibilidade de dados robustos em estações vizinhas ou próximas à qual pretende-se preencher tais falhas é condição rara nos repositórios de dados meteorológico-climáticos.

Que os resultados obtidos para este estudo se aplicam apenas a ele. Entretanto, as técnicas aqui demonstradas, além de serem de fácil aplicação, são as mais amplamente empregadas nos mais diversos trabalhos que versam sobre o tema – preenchimento de falhas em dados pluviométricos.

Por fim, para além dos resultados (o preenchimento das falhas) e das técnicas utilizadas como ferramentas ao seu preenchimento, atenção especial deve ser dada a avaliação da qualidade/eficácia dos dados preenchidos, considerando as simulações feitas (e avaliadas) para anos classificados como Habituais, Secos e Chuvosos.

Isso demonstra, entre outros aspectos, que as análises desenvolvidas em Climatologia, seja Geográfica ou Tradicional, podem (e devem) se apropriar de técnicas estatísticas consagradas, minimamente, na preparação de seus dados para tratamento/análises mais profundas e complexas.

E por fim, mas não encerrando o assunto, embora aparentemente tais linhas metodológicas e de abordagem sejam antagônicas, elas podem ser (e são), na verdade, complementares àqueles que fazem uso de suas técnicas em seus objetos de estudo.

REFERÊNCIAS

AB'SÁBER, A. **Os domínios de natureza do Brasil**. São Paulo: Ateliê Editorial, 2003.

AMADOR, M. C.; SILVA, C. A. DA. O preenchimento de falhas de séries históricas de precipitação. **Revista Eletrônica da Associação dos Geógrafos Brasileiros, Seção Três Lagoas**, v. 1, n. 31, p. 178-202, 1 jun. 2020.

ARAÚJO, S.R.; SANCHES, F.O.; FERREIRA, R.V. Metodologia para validação de dados TRMM para Uberaba (MG). **Entre Lugar**, v. 8, p. 131-141, 2017.

BERTONI, J.C.; TUCCI, C.E.M. Precipitação. In: TUCCI, C.E.M. **Hidrologia: ciência e aplicação**. Porto Alegre: UFRGS, 2007.

CABRAL JÚNIOR, J.B, SILVA, C.M.S.E.; DE ALMEIDA, H.A. BEZERRA, B.G.; SPYRIDES, M.H.C. Detectando tendência linear da evapotranspiração de referência em áreas de cultivo irrigado no semiárido brasileiro. **Theor Appl Climatol,** 138**,** 215–225 2019.

CABRAL JÚNIOR, J.B.; LUCENA, R.L. Analysis of precipitations by non-parametric tests of Mann-Kendall and Kruskal-Wallis. **Mercator**, Fortaleza, v. 19, dez. 2019.

DIAZ, C.C.F.; PEREIRA, J.A.S.; NÓBREGA, R.S. Comparação de dados estimados por dois diferentes métodos para o preenchimento de falhas de precipitação pluvial na bacia do rio Pajeú, Pernambuco, Brasil. **Revista Brasileira de Climatologia**, v.22, jan/jun., 2018.

EDUARDO, C. C. **Cartografia geomorfológica comparada: aplicações no município de Juiz de Fora-MG como subsídio ao planejamento**. Dissertação de Mestrado (em Geografia), Instituto de Ciências Humanas- ICH, Universidade Federal de Juiz de Fora, Juiz de Fora, 2018.

FERREIRA, C.C.M. Estudo do comportamento do período chuvoso em Juiz de Fora-MG. **Revista Geonorte**. UFAM, v.1, n.5, 2012, p. 953-963.

GALVANI, E.; LUCHIARI, A. Critérios para classificação de anos com regime pluviométrico normal, seco e úmido. In: GALVANI, E.; LIMA, N.G.B. (Org.). **Climatologia Aplicada**: Resgate aos estudos de caso. Curitiba: Editora CRV, 2012.

HINKLE D.E.; WIERSMA W.; JURS, S.G. **Applied Statistics for the Behavioral Sciences**. 5th ed. Boston: Houghton Mifflin; 2003.

JUNQUEIRA, R.; AMORIM, J.S.; OLIVEIRA, A.S. Comparação entre diferentes metodologias para preenchimento de falhas em dados pluviométricos. **Sustentare**, v.2, n.1, jan/jul., 2018.

MELLO, Y.R.; KOHLS, W.; OLIVEIRA, T.M. Uso de diferentes métodos para o preenchimento de falhas em estações pluviométricas. **Boletim de Geografia**, Maringá, v.35, n.1, p.112-121, 2017.

MILANESI, M.A.; GALVANI, E. Preenchimento de falhas em banco de dados de curta série de observações. In: **Anais...** Simpósio Brasileiro de Climatologia Geográfica, Rondonópolis: UFMT, 2006, v.1, p. 1-10.

NASCIMENTO, T.S.; SARAIVA, J.M.; SENNA, R.; AGUIAR, F.E.O. Preenchimento de falhas em banco de dados pluviométricos com base em dados do CPC (Climate Prediction Center): estudo de caso do rio Solimões-Amazonas. **Revista Brasileira de Climatologia**, v.7, set., 2009.

OLIVEIRA, L.F.C.; FIOREZE, A.P.; MEDEIROS, A.M.M.; SILVA, M.A.S. Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.14, n.11, 2010.

OLIVEIRA, T.A.; TAVARES, C.M.G.; SANCHES, F.; FERREIRA, C.C.M. Variabilidade pluviométrica no município de Juiz de Fora-MG no período de 1910-2018: investigação a partir da técnica do box plot. **Revista Brasileira de Climatologia**, v. 26, p. 457-478, 2020.

PECHUTTI, K.F.; SANT'ANNA NETO, J.L. Técnicas estatísticas para a homogeneização de dados de temperatura em séries temporais climatológicas. **Revista Brasileira de Climatologia**, v.18, jan/jun., 2016.

QUINA, R.R.; COSTA, R.M.; SANCHES, F.O.; GOMES, F.C.M.; ZAIDAN, R.T. Comparação de produtos altimétricos dos sensores Palsar/Alos e Lidar, a partir do modelo Sinmap. **Research, Society and Development**, v. 9, p. e27391211149-26, 2020.

RUEZZENE, C.B.; MIRANDA, R.B.; TECH, A.R.B.; MAUAD, F.F. Revisão bibliográfica dos principais métodos de preenchimento de falhas em dados de precipitação. **Brazilian Journals of Business**, Curitiba, v. 2, n. 3, p. 2057-2071, jul./set. 2020.

SANCHES, F.O.; BALEN, D.S.; SILVA, R.V.; ROSA, K.K.; RADÜNZ, A. Chuvas no Rio Grande do Sul: um estudo sobre as precipitações acumuladas intensas no Alto Uruguai gaúcho. **Revista Brasileira de Climatologia**, v.15, jul/dez., 2014a.

SANCHES, F.O.; VERDUM, R.; FISCH, G. Tendência de longo prazo das chuvas diárias no Sudoeste do Rio Grande do Sul: os eventos extremos e a arenização. **Revista Brasileira de Geografia Física**, v.7, n.6, p.1100-1109, 2014b.

SPIEGEL, M.R. **Estatística**. 3.ed. São Paulo: Pearson Makron Books/McGraw-Hill do Brasil, 1993.

VENTURA, T.M.; SANTANA, L.L.R.; MARTINS, C.A.; FIGUEIREDO, J.M. Análise da aplicabilidade de métodos estatísticos para preenchimento de falhas em dados meteorológicos. **Revista Brasileira de Climatologia**, v.19, jul/dez., 2016.

Recebido em setembro de 2021.

Revisão realizada em outubro de 2021.

Aceito para publicação em dezembro de 2021.