

# **Comparação entre métodos de preenchimento de falhas em séries de dados meteorológicos da bacia hidrográfica do Rio das Velhas (MG)**

Lilian Aline Machado  
Universidade Federal de Minas Gerais – UFMG  
Wellington Lopes Assis  
Universidade Federal de Minas Gerais – UFMG

## **Resumo**

Este trabalho tem como objetivo a comparação de diferentes métodos estatísticos para o preenchimento de falhas em séries de dados meteorológicos de estações existentes na bacia hidrográfica do Rio das Velhas. Foram testados dez diferentes métodos estatísticos de preenchimento de falhas para as variáveis precipitação, temperatura média, temperatura máxima, temperatura mínima e umidade relativa do ar. A série de dados originais de dez anos (2003-2012) foi falhada aleatoriamente em 10% e 30% para a aplicação dos testes e comparação dos dados simulados e observados. Para verificar os resultados obtidos foram utilizados o erro médio, o coeficiente de correlação, o desvio padrão e o teste de dupla massa. O método de ponderação regional apresentou resultados ligeiramente melhores para todas as variáveis, embora outros métodos como a ponderação regional com base em regressões lineares e a regressão múltipla também tenham apresentado bons resultados.

**Palavras-chave:** Dupla massa; ponderação regional; Rio das Velhas

## **Abstract**

This paper aims to compare different statistical methods for filling in gaps in meteorological data series of stations in the Rio das Velhas watershed. Ten different statistical methods of fault filling were tested for the variables precipitation, average temperature, maximum temperature, minimum temperature and relative humidity. The original ten-year data series (2003-2012) had a random 10% and 30% failure to apply the test and to compare simulated and observed data. To verify the results obtained, the mean error, the correlation coefficient, the standard deviation and the double mass test were used. The regional weighting method showed slightly better results for all variables, although other methods such as regional weighting based on linear regressions and multiple regression also showed good results.

**Keywords:** Doble mass; regional weighting, Rio das Velhas

lilian.aline.machado@hotmail.com  
assisw@gmail.com

## Introdução

Os dados meteorológicos mensurados por estações de superfície são de grande importância científica, uma vez que o resultado das análises climáticas pode auxiliar na tomada de decisões em diferentes áreas do planejamento territorial. No Brasil há uma deficiência em relação a séries históricas sequenciais de dados meteorológicos, seja esse pontual, apresentado falhas de dados dentro das medições de um mesmo posto, ou pela cobertura heterogênea do território nacional, com lacunas significativas de estações em áreas de grande interesse científico, econômico e social, como a Amazônia.

Segundo Diaz *et al.* (2018) as falhas ocorrem quando a coleta feita de forma convencional, realizada por meio da anotação dos dados por um indivíduo coletor, é deficitária em algum momento, ou por erros e sobreposição de dados na coleta em estações automáticas. A presença de falhas ao longo de uma série histórica de dados pode levar a resultados equivocados ou tendenciosos, comprometendo a análise climática e, conseqüentemente, podendo prejudicar o planejamento territorial (VIEIRA *et al.*, 2018). A escassez dos dados e as múltiplas lacunas nas séries existentes exigem dos pesquisadores que necessitam de séries completas a adoção de técnicas e metodologias de preenchimento dos dados ausentes baseadas em geo-estatística.

No Brasil, a principal fonte de dados oriundos de estações meteorológicas é administrada e disponibilizada pelo Instituto Nacional de Meteorologia (INMET). As séries históricas de dados meteorológicos disponibilizados pelo INMET<sup>1</sup> normalmente apresentam falhas em escala temporal horária, diária ou mensal. Ainda que a disponibilidade *on line* e o *download* gratuito dessas informações sejam um avanço para o campo do ensino e pesquisa da climatologia brasileira, já que até meados dos anos 2000 esses dados eram de difícil acesso, as falhas ainda consistem em um limite para o desenvolvimento das pesquisas no âmbito da Geografia e das Ciências Ambientais.

Considerando-se o exposto, esse artigo tem como objetivo a comparação de diferentes métodos estatísticos para o preenchimento de falhas em séries de dados meteorológicos históricos de estações existentes na bacia hidrográfica do Rio das Velhas. As séries foram falhadas aleatoriamente no intuito de se comparar estatisticamente os resultados simulados com os dados observados. O melhor método observado para cada variável meteorológica analisada foi empregada para preencher as falhas das estações meteorológicas no interior e entorno da bacia no intuito de subsidiar o mapeamento de unidades climáticas dessa pesquisa de doutorado realizada no programa de Pós-Graduação em Geografia da UFMG.

## Fundamentação teórica e bases conceituais

Os fenômenos meteorológicos se estabelecem a partir da interação entre elementos climáticos e os fatores geográficos, sendo a resposta atmosférica a esse processo tão complexa quanto é a composição geoecológica da superfície (VENTURA *et al.*, 2013). Portanto, para preencher dados ausentes é necessário estimar valores a partir da modelagem do comportamento atmosférico tendo como base dados históricos.

<sup>1</sup>Disponível de forma gratuita através do Banco de Dados Meteorológicos Para Ensino e Pesquisa (BDMEP) em <<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>>.

A estimação de dados meteorológicos falhos pode ser feita utilizando-se os próprios dados da estação meteorológica alvo do preenchimento, dados entre estações ou através de métodos regressivos (ALLEN E DEGAETANO, 2001). No primeiro caso, estações com percentual de falhas pequeno, por exemplo, abaixo de 5%, adquirem bons resultados com essa modalidade de métodos (KEMP *et al.*, 1983; ACOCK E PACHEPSKY, 2000). Já quando aumenta o percentual de falhas, o uso de estações vizinhas ou métodos regressivos pode trazer melhores resultados para a simulação dos dados ausentes. No entanto, o método mais apropriado dependerá de uma série de fatores como o objetivo de uso, o percentual de falhas em relação à série histórica, e a disponibilidade de dados em estações meteorológicas vizinhas para o uso no preenchimento de falhas (KEMP *et al.*, 1983).

Dentre os métodos mais simples está a média aritmética. Essa consiste na soma do valor anterior e posterior à falha dividida por dois (Equação 1). Quando o valor que será utilizado para o cálculo da média também é ausente, ele é ignorado e o anterior ou posterior é utilizado. A média simples é indicada para preencher dados de precipitação quando a porcentagem de valores perdidos é inferior a 10% da série histórica (TEEGAVARAPU e CHANDRAMOULI, 2005).

$$X = \frac{x_1 + x_2}{2} \quad \text{Equação 1}$$

Onde  $x$  é a variável simulada,  $x_1$  é o valor anterior à falha e  $x_2$  é o valor posterior à falha.

Ventura *et al.* (2016) encontraram um bom ajuste da média simples para o preenchimento de falhas horárias da pressão atmosférica. No entanto, para as variáveis temperatura, umidade relativa do ar e ponto de orvalho, embora os dados simulados respondessem a mais de 80% dos dados reais, tiveram melhor ajuste com o método de regressão múltipla. Vieira *et al.* (2018) também encontraram bom ajuste da média simples para preenchimento de falhas mensais de temperatura máxima e precipitação. Porém, como houve falha sequencial ao longo de três anos da série utilizada pelos autores, resultando na repetição de valores preenchidos pela média simples, esse método foi descartado. Para variáveis de distribuição pontual, como a precipitação, a média pode não ser um bom método de preenchimento como observou Tung (1983).

A média também pode ser utilizada considerando-se uma série de dados anteriores à falha como forma de reduzir a influência de valores extremos sobre o resultado do cálculo (Equação 2). Esse método, denominado média móvel, consiste no cálculo da média aritmética de três ou mais dados anteriores à falha.

$$X = \frac{\sum X_n}{n} \quad \text{Equação 2}$$

Onde  $x$  é a variável simulada,  $x_n$  são os valores anteriores à falha utilizados e  $n$  corresponde à quantidade de valores anteriores à falha utilizados.

Ventura *et al.* (2016) não encontraram bons resultados para a aplicação da média móvel no preenchimento de falhas horárias da temperatura, umidade, ponto de orvalho e pressão, observando índices de correlação abaixo de 0,8.

Ainda no âmbito dos métodos que utilizam dados da própria estação meteorológica para o preenchimento de falhas, a análise de regressão é realizada de forma a definir as correlações entre duas ou mais variáveis que apresentem em sua natureza relações de causa e efeito. Na regressão simples, correlaciona-se duas variáveis com maior relação de causa e efeito entre si, enquanto na regressão múltipla correlaciona-se a variável dependente com duas ou mais variáveis para a simulação. Assim, é estabelecida a correlação entre variáveis meteorológicas da série histórica do posto com falhas. As variáveis que apresentam melhor correlação compõem a equação de regressão (Equação 3).

$$y = \alpha + \beta x \quad \text{Equação 3}$$

Onde  $y$  a variável dependente,  $\alpha$  é o coeficiente linear,  $\beta$  o coeficiente angular e  $x$  a variável independente.

Ventura *et al.* (2016) testaram o preenchimento de falhas geradas aleatoriamente em uma série de dados e encontraram melhores resultados de simulação para as variáveis temperatura, umidade relativa do ar e ponto de orvalho com a regressão linear múltipla, em comparação à regressão simples, média simples e média móvel. Mello *et al.* (2017) encontrou bons resultados de preenchimento de dados com o método de regressão linear múltipla para a variável precipitação em relação aos métodos de ponderação regional e ponderação regional com base em regressões lineares. Yaguchi *et al.* (2016) também encontrou bons resultados para a regressão linear múltipla ao preencher dados diários de temperatura máxima e mínima do ar.

Enquanto os métodos apresentados até aqui estão baseados nos dados da própria estação com falhas para o preenchimento, outros consideram o comportamento dos dados meteorológicos em postos vizinhos, em uma área cuja repercussão climática regional siga padrões semelhantes. Dentre eles, há a ponderação regional. Neste método são selecionados pelo menos três postos que possuam no mínimo 10 anos de dados e que se localizem em uma região com características climáticas semelhantes à estação meteorológica que apresenta falhas (OLIVEIRA *et al.*, 2010). Então, é calculada a soma da razão entre a variável correspondente à data da falha da estação base e a média dessa variável de cada estação, e em seguida este valor é multiplicado pela razão entre a média da variável da estação base e o número de estações de referência utilizadas (Eq. 4).

$$x = \frac{1}{3} \left( \frac{x_1}{xm_1} + \frac{x_2}{xm_2} + \frac{x_3}{xm_3} \right) ym \quad \text{Equação 4}$$

Onde  $x$  é a falha a ser preenchida;  $x_1, x_2, x_3$  são as variáveis correspondentes ao ano que se deseja preencher, observadas em 3 estações vizinhas;  $y_m$  é a variável média do posto  $y$ ; e  $x_{m1}, x_{m2}, x_{m3}$  são as variáveis médias das 3 estações circunvizinhas.

A ponderação regional pode ser aprimorada quando utilizada junto à regressão, associando a média das variáveis simuladas à correlação entre estações meteorológicas vizinhas. Este método consiste em estabelecer regressões lineares entre o posto com dados a serem preenchidos,  $Y$ , e cada um dos postos vizinhos,  $X_1, X_2, \dots, X_n$ . De cada uma das regressões lineares calculadas obtém-se o coeficiente de correlação  $r$ . Este coeficiente é multiplicado pela variável em cada posto de referência, e os valores são somados e divididos pela soma dos coeficientes de correlação dos postos de referência (Equação 5).

$$y = \frac{((x_1 r_1) + \dots + (x_n r_n))}{(r_1 + \dots + r_n)} \quad \text{Equação 5}$$

Onde  $y$  é a falha a ser preenchida,  $x$  é a variável de cada estação meteorológica vizinha de referência e  $r$  é o índice de correlação entre as variáveis da estação de referência e a estação base.

Mello *et al.* (2017) obtiveram bons resultados para o preenchimento de falhas de precipitação mensal para o município de Joinville utilizando a ponderação regional baseada em regressões lineares (SC). No entanto, os autores encontraram resultados semelhantes para a regressão linear múltipla e a ponderação regional com base em regressões lineares, destacando que a melhor opção seria a utilização de diferentes métodos escolhendo o melhor para cada estação. Oliveira *et al.* (2010) também encontraram bons resultados com a ponderação regional, junto à regressão linear múltipla, para a região central do estado de Goiás.

O método Inverso do Quadrado da Distância (Inverse Distance Weight – IDW) é amplamente utilizado para preenchimento de falhas de dados de precipitação (TEEGAVARAPU e CHANDRAMOULI, 2005). Para esse método, é realizada a média ponderada pelo inverso da distância entre a localização da estação base com dados ausentes e estações de referência vizinhas a ela (Equação 6).

$$y = \frac{\left(\frac{x_1}{d_1} + \dots + \frac{x_n}{d_n}\right)}{\left(\frac{1}{d_1} + \dots + \frac{1}{d_n}\right)} \quad \text{Equação 6}$$

Onde  $y$  é a falha a ser preenchida,  $x$  é a variável de cada estação meteorológica vizinha a estação com o dado ausente, e  $d$  é a distância em metros da estação vizinha àquela com o dado ausente.

A seleção do número de estações vizinhas àquela em que se fará o

preenchimento das falhas deve considerar de três ou quatro estações, embora o número exato seja arbitrário, segundo Teegavarapu e Chandramouli (2005). Os resultados do preenchimento serão influenciados pelo número de estações e pela distância entre elas. Embora seja comum o uso do IDW para simular a precipitação, Tung (1983) não encontrou bons resultados desse método para simulação da variável em regiões montanhosas.

Entre os métodos que consideram o comportamento de estações vizinhas àquela com falhas, está a técnica utilizada pelo U.K. Meteorological Office. Segundo Tabony (1983), uma estação vizinha que apresente uma boa correlação com os dados da estação base deve ser utilizada como referência para o preenchimento das falhas. A estimativa é feita a partir do cálculo da diferença entre o valor médio mensal da variável meteorológica entre a estação base e a de referência. Para as variáveis contínuas, como temperatura e umidade relativa do ar preenche-se o dado ausente somando-se a diferença média da variável entre as estações. Para a precipitação é assumida uma razão constante baseada na diferença média de precipitação entre as estações. Bier e Ferraz (2017), ao preencher falhas de dados de temperatura para o estado do Rio Grande do Sul, encontraram os melhores resultados com os métodos UK, ponderação regional e regressão linear múltipla.

Enquanto os métodos supracitados consideram a similaridade temporal dos dados e a proximidade geográfica das estações meteorológicas, outros consideram o efeito dos fatores geográficos na simulação de dados para o preenchimento de falhas. Este é o caso da proposta de Pinto e Alfonsi (1974). Os autores propuseram um meio de cálculo que considera a altitude e latitude para simular dados ausentes de temperaturas mínima, média e máxima. O cálculo envolve o estabelecimento de uma equação de regressão entre os valores de temperatura da série histórica, altitude (em metros) e latitude (em minutos).

Para o cálculo dos valores ausentes no método de Pinto e Alfonsi, é determinada a média da temperatura do ar mínima, média e máxima mensal do período que abrange a série de dados da estação. Em seguida é estabelecida a equação de regressão linear múltipla mensal entre temperatura, altitude e latitude, considerando-se os valores médios históricos da série de temperatura. A equação de regressão definida é então utilizada para preencher os dados faltantes.

Fante e Sant'Anna Neto (2016) testaram a técnica de Pinto e Alfonsi em comparação à regressão linear e ponderação regional para o preenchimento de dados de temperatura de estações meteorológicas do estado de São Paulo. Embora os melhores resultados tenham sido obtidos com o método de ponderação regional, a técnica de Pinto e Alfonsi apresentou bons resultados.

Vale ressaltar que após aplicados os métodos de preenchimento de falhas, é necessário atestar a consistência da série de dados meteorológicos. Dentre as técnicas mais comuns está o teste de dupla massa, o qual consiste em um método desenvolvido pelo Serviço Geológico dos Estados Unidos para atestar a consistência de séries de dados meteorológicos mensais e anuais. A metodologia consiste em “selecionar os postos de uma região, acumular para cada um deles os valores mensais, se for o caso, e plotar num gráfico cartesiano os valores acumulados correspondentes ao posto a validar (nas ordenadas) e de um outro posto confiável adotado como base de comparação (nas abscissas)” (ANA, 2011, p. 12).

Para que haja consistência dos dados, deve haver uma tendência linear em relação à estação vizinha (OLIVEIRA *et al.*, 2010). Se houver alteração significativa, afirma-se que os dados não foram corretamente medidos ou são hidrológicamente diferentes (MELLO; SILVA, 2013). Por esse método é possível identificar erros sistemáticos (mudança de declividade ou tendência), erros de transcrição ou postos sujeitos a diferentes regimes pluviométricos.

<sup>2</sup>Considerando-se a média das falhas (22%) diminuída do desvio padrão (12%).

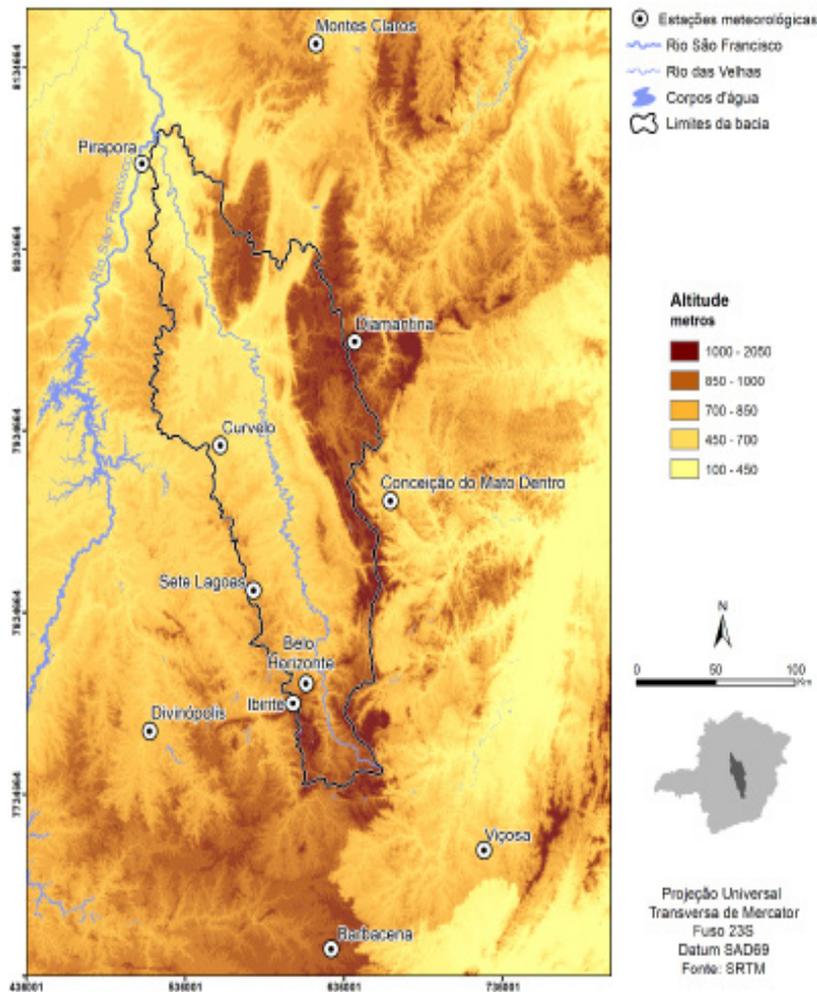
<sup>3</sup>Considerando-se um valor próximo da média das falhas (22%) acrescida do desvio padrão (12%).

## Procedimentos metodológicos

Para atingir os objetivos propostos de estudo, foram obtidos dados mensais das variáveis meteorológicas temperatura (média compensada, máxima e mínima), umidade relativa do ar e precipitação de onze estações meteorológicas pertencentes ao Instituto Nacional de Meteorologia (INMET). As estações meteorológicas escolhidas encontram-se dentro dos limites da bacia hidrográfica do Rio das Velhas e em seu entorno em uma distância máxima de 100 km (Fig. 1).

Após obtidos os dados, foram contabilizados o número de meses com ausência de informações no período de 1988 a 2017, período de 30 anos mais próximo da atualidade, e calculada a porcentagem de falhas em cada série temporal (Fig. 2). O ano de 2018 foi desconsiderado por não contemplar dados para sete das onze estações meteorológicas selecionadas. A média de falhas entre as estações foi de 22% dos dados da série histórica, com um desvio padrão de 12%. Assim, foi definido que o teste seria realizado sobre uma série de dados aleatoriamente falhados em 10%<sup>2</sup> e 30%<sup>3</sup>.

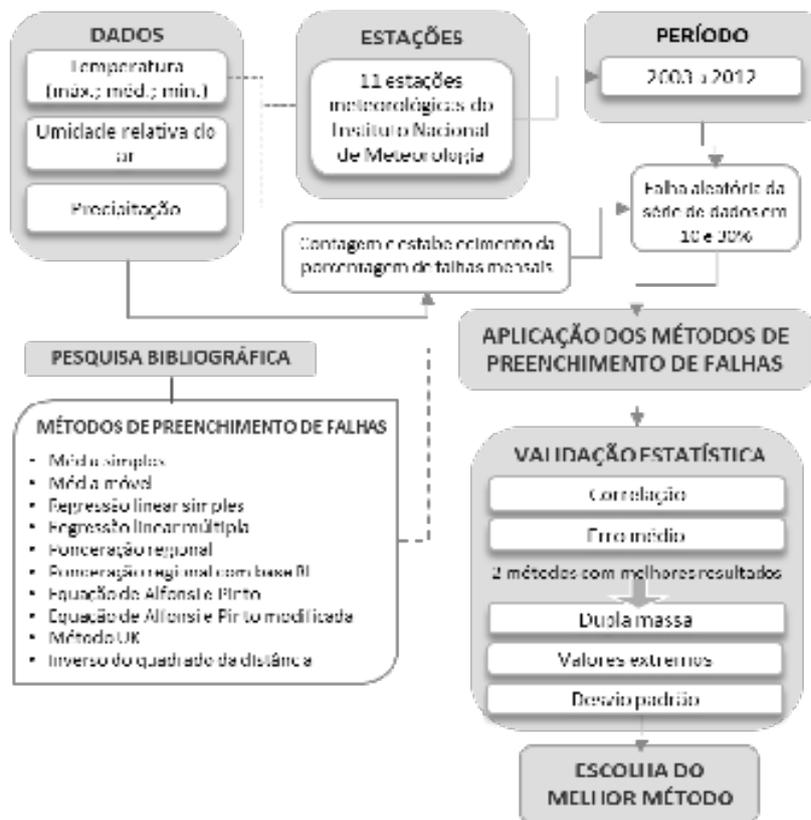
Foi definido o período de 2003 a 2012 (dez anos) para realização dos testes em função desse período apresentar o menor percentual de falhas entre as estações meteorológicas ao longo dos anos, um valor médio abaixo de 5%. Em seguida foi feita a falha aleatória de dados no período de 2003 a 2012 com 10% e 30% de ausência nos registros a fim de serem aplicados os métodos de preenchimento de falhas.



Elaboração: Autores.

**Figura 1: Localização das estações meteorológicas convencionais do INMET.**

Os métodos de preenchimento foram escolhidos após realizada uma revisão bibliográfica em revistas nacionais e internacionais. Foram escolhidos métodos que utilizam dados da própria estação, de estações vizinhas e dados ambientais locais para o preenchimento das falhas, sendo eles: média simples e média móvel (VENTURA *et al.*, 2016); regressão simples e regressão múltipla (MELLO *et al.*, 2017); ponderação regional e ponderação regional com base em regressões lineares (OLIVEIRA *et al.*, 2010), inverso do quadrado da distância (TEGAVARAPU e CHANDRAMOULI, 2005), método do UK Meteorological Office (TABONY, 1983) e; equações de Alfonsi e Pinto (FANTE e SANT'ANNA NETO, 2016).



**Figura 2: Roteiro metodológico. Elaboração: Autores.**

O método proposto por Alfonsi e Pinto, entretanto, gerou resultados repetitivos uma vez que considerada a média mensal dos dados meteorológicos do período analisado, junto aos dados de altitude e latitude, para simular o dado ausente. A equação de regressão entre estes dados foi, então, utilizada para preencher os dados faltantes. O valor simulado acabou sendo o mesmo para cada mês, ou seja, todas as falhas observadas em janeiro, por exemplo, tiveram o mesmo valor simulado. Ciente da variabilidade anual inerente ao clima, foi feita uma alteração no método incluindo-se no cálculo da regressão os dados de temperatura média mensal (meses individuais e não a média da série histórica) para simulação da temperatura mínima e máxima, e temperatura mínima para simulação dos dados de temperatura média. Assim foram testados o método tradicional de Alfonsi e Pinto e o método com essa modificação.

Em seguida, os valores simulados através dos métodos e os valores observados foram comparados estatisticamente utilizando-se o coeficiente de determinação ( $R^2$ ), o qual indica quanto da variação total é comum aos elementos que constituem os pares analisados, como uma maneira de se verificar se o modelo matemático proposto é adequado ou não para descrever o fenômeno. Também foi calculado o erro médio das séries de dados simulados, em comparação à observadas. Os dois métodos que apresentaram maiores valores de  $R^2$  e menores erros, foram selecionados para que fosse testada a consistência dos dados simulados em comparação com os observados através do teste de dupla massa.

Foi realizada ainda a comparação dos valores mínimos, máximos e de desvio padrão entre as séries de dados observados e simulados dos métodos que apresentaram melhores resultados. Através dessa etapa, concluiu-se a escolha dos métodos mais adequados para preencher as falhas da série para cada

variável meteorológica.

## Resultados

A variável precipitação não apresentou índices de determinação entre os valores observados e simulados acima de 0,9 (Tab. 1). A melhor correlação obtida para 10% de falhas foi por meio dos métodos ponderação regional (0,81) e ponderação regional RL (0,81). Para 30% de falhas, o primeiro método foi o único que apresentou coeficiente de correlação acima de 0,8.

**Tabela 1: Coeficiente de correlação entre valores observados e simulados de acordo com cada método de preenchimento de falhas**

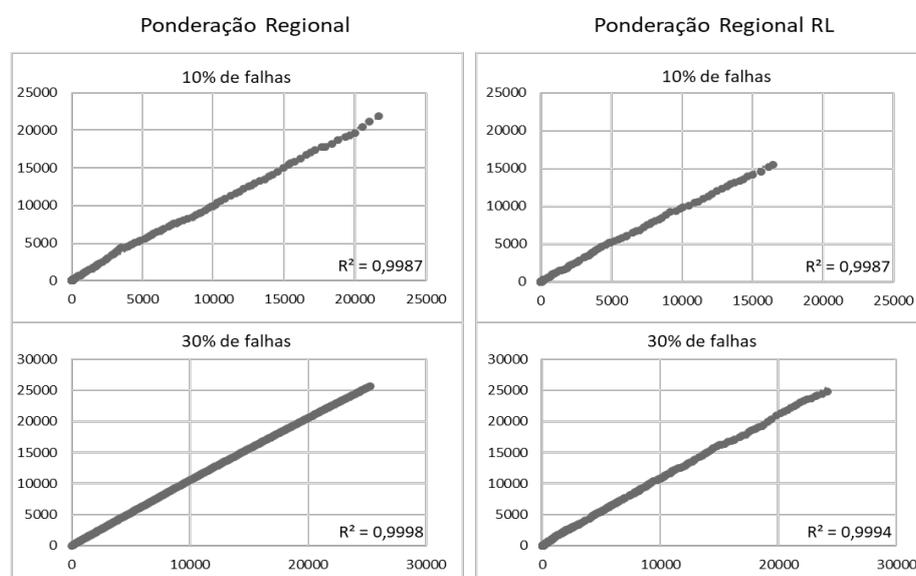
Falhas	Precipitação		Temperatura Média		Temperatura Máxima		Temperatura Mínima		Umidade Relativa do Ar	
	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
Média Simples	0,31	0,79	0,86	0,88	0,69	0,81	0,88	0,88	0,59	0,81
Média Móvel	0,06	0,66	0,36	0,83	0,40	0,85	0,27	0,79	0,18	0,77
Regressão Simples	0,19	0,36	0,85	0,83	0,00	0,81	0,85	0,83	0,16	0,20
Regressão Múltipla	0,49	0,48	0,98	0,98	0,96	0,96	0,98	0,96	0,25	0,23
Ponderação Regional	0,81	0,83	0,96	0,96	0,96	0,98	0,96	0,98	0,94	0,92
Ponderação Regional RL	0,81	0,79	0,81	0,77	0,69	0,48	0,77	0,69	0,69	0,58
IDW	0,58	0,69	0,77	0,76	0,61	0,58	0,67	0,64	0,64	0,64
Método UK	0,74	0,69	0,90	0,92	0,90	0,92	0,92	0,92	0,79	0,79
Equações de Alfonsi	-	-	0,85	0,85	0,77	0,71	0,79	0,85	-	-
Equações de Alfonsi Mod.	-	-	0,92	0,92	0,66	0,90	0,64	0,74	-	-

A ponderação regional (PR) e a ponderação regional RL (PRRL) também apresentaram os menores erros médios para as porcentagens de falhas adotadas no teste (Tab. 2). Para 10% de falhas, a PR apresentou 1,3mm e a PRRL, 0,3mm de erro médio. Para 30% de falha, o menor erro médio foi observado no método UK (0,05mm), enquanto a PR apresentou 2,82mm e a PRRL, 2,57mm.

**Tabela 2: Erro médio dos valores simulados de acordo com cada método de preenchimento de falhas**

Falhas	Precipitação		Temperatura Média		Temperatura Máxima		Temperatura Mínima		Umidade Relativa do Ar	
	10%	30%	10%	30%	10%	30%	10%	30%	10%	30%
	Média Simples	19,3	-3,6	-0,09	-0,07	-0,21	-0,1	0,09	-0,09	1,15
Média Móvel	26,2	-3,33	-0,09	0,01	-3,22	0,03	0,19	-0,03	2,36	-0,26
Regressão Simples	1,60	15,52	0,03	-0,13	0,20	2,46	0,01	0,17	0,14	-0,24
Regressão Múltipla	11,6	11,6	0,01	0,02	-0,01	-0,01	0,01	-0,02	0,01	-0,04
Ponderação Regional	1,30	2,82	-0,04	0,73	0,90	-0,05	0,02	-0,03	-0,04	0,34
Ponderação Regional RL	0,30	2,57	-0,02	-0,22	-0,03	-0,33	-0,06	-0,65	0,47	4,75
IDW	10,7	-3,93	-0,35	0,31	-0,52	0,5	-0,29	0,24	-1,22	0,96
Método UK	11,3	0,05	0,13	-0,03	0,22	-0,22	-0,11	0,07	-0,88	0,98
Equações de Alfonsi	-	-	0,07	-0,03	0,00	-0,02	0,05	-0,02	-	-
Equações de Alfonsi Mod.	-	-	0,12	-0,02	0,67	-0,01	-0,72	0,46	-	-

Os resultados do teste de dupla massa apontaram que ambos os métodos, ponderação regional e ponderação com base em regressões lineares, apresentam boa consistência dos dados, embora o primeiro tenha apresentado valores mais alinhados com a reta de tendência (Fig. 3).



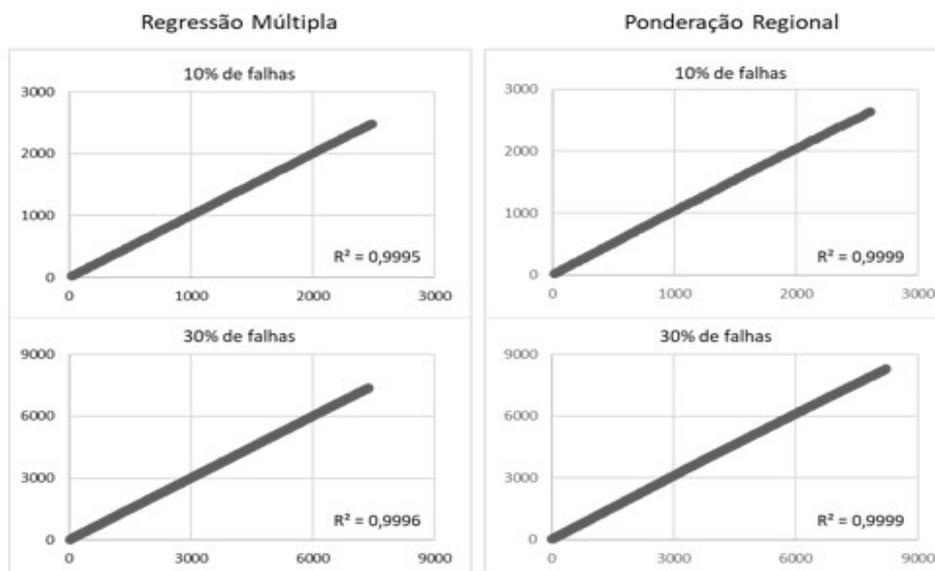
**Figura 3: Teste de dupla massa para os métodos de ponderação regional e ponderação regional com base em regressões lineares aplicado à variável precipitação.**

Quanto à amplitude dos dados simulados e observados, a ponderação com base em regressões lineares apresentou uma amplitude maior entre os valores simulados e observados, assim como valores mais distantes do desvio padrão das amostras observada e simulada do que a regressão linear (Tab. 3).

**Tabela 3: Valores mínimos, máximos e desvio padrão dos dados de precipitação simulados e observados**

		10%		30%	
		Simulado	Observado	Simulado	Observado
Ponderação Regional	Mínimo	0	0	0	0
	Máximo	626,5	720,0	626,5	725,0
	Desvio Padrão	167,7	173,5	119,4	123,0
Ponderação Regional RL	Mínimo	0	0	0	0
	Máximo	604,5	720,0	636,5	725,0
	Desvio Padrão	163,8	138,1	132,8	147,0

Os dados de temperatura média do ar apresentaram coeficiente de determinação acima de 0,96 nos métodos de regressão múltipla e ponderação regional para 10% e 30% de falhas. Para ambos o erro médio esteve entre -0,13 e 0,17. Os métodos apresentaram boa consistência de acordo com o teste de dupla massa, porém a regressão múltipla apresentou correlações um pouco melhores nesse teste (Fig. 4).



**Figura 4: Teste de dupla massa para os métodos de ponderação regional e regressão múltipla com base em regressões lineares aplicado à variável temperatura média do ar.**

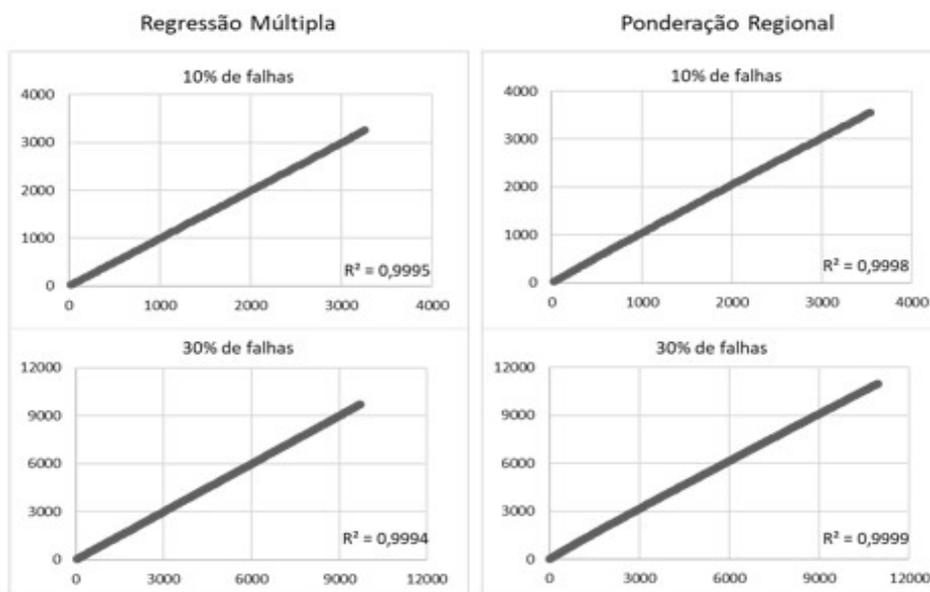
A diferença entre os valores máximos e mínimos simulados e observados foi de 0,1°C para a ponderação regional, enquanto na regressão múltipla esse valor foi de até 1,0°C (Tab. 4). A diferença de valores do desvio padrão entre as amostras simuladas e observadas foi de -0,1°C nas simulações com 10% de falhas para a ponderação regional, e -0,5°C para a regressão múltipla. Já para 30% de falhas o desvio padrão das amostras foram iguais para a ponderação regional e houve uma diferença de -0,3°C para a regressão múltipla.

**Tabela 4: Valores mínimos, máximos e desvio padrão dos dados de temperatura média simulados e observados**

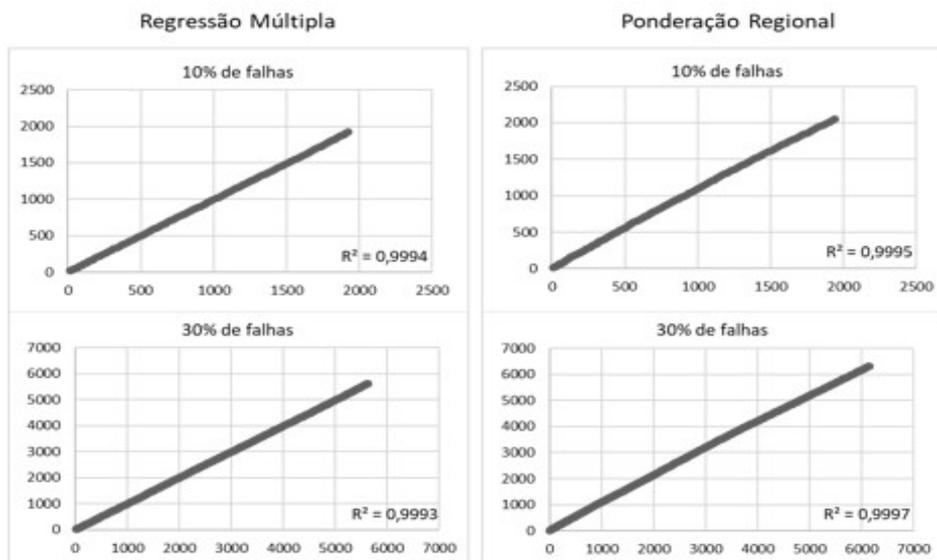
		10%		30%	
		Simulado	Observado	Simulado	Observado
Ponderação Regional	Mínimo	14,5	14,4	14,5	14,4
	Máximo	27,2	27,2	28,1	28,0
	Desvio Padrão	3,1	3,2	3,1	3,1
Regressão Múltipla	Mínimo	14,4	14,4	14,1	14,4
	Máximo	26,3	27,2	27,8	28,0
	Desvio Padrão	2,7	3,2	2,8	3,1

Para a temperatura máxima e mínima, as melhores correlações (0,96 e 0,98) também foram encontradas com os métodos ponderação regional e regressão múltipla em ambos os percentuais de falhas. O erro médio foi mais baixo nas simulações com o método de ponderação regional, para a temperatura máxima, enquanto para a temperatura mínima a diferença entre os métodos não ultrapassou 0,01°C.

A consistência dos dados teve boa correlação para esses métodos, embora a ponderação regional tenha apresentado um resultado levemente melhor do que a regressão múltipla tanto para a temperatura máxima (Fig. 5), quanto para a mínima (Fig. 6).



**Figura 5: Teste de dupla massa para os métodos de ponderação regional e ponderação regional com base em regressões lineares aplicado à variável temperatura máxima do ar.**



**Figura 6: Teste de dupla massa para os métodos de ponderação regional e ponderação regional com base em regressões lineares aplicado à variável temperatura mínima do ar.**

Com relação à simulação dos valores extremos da temperatura máxima, a ponderação regional apresentou uma diferença máxima entre valores simulados e observados de 0,8°C para a ponderação regional e até 1,0°C para a regressão múltipla (Tab. 5).

**Tabela 5: Valores mínimos, máximos e desvio padrão dos dados de temperatura máxima simulados e observados**

		10%		30%	
		Simulado	Observado	Simulado	Observado
Ponderação Regional	Mínimo	23,0	23,3	21,2	21,3
	Máximo	27,2	27,2	33,9	33,3
	Desvio Padrão	2,0	2,0	2,8	2,8
Regressão Múltipla	Mínimo	24,3	23,3	20,5	21,3
	Máximo	26,2	27,2	34,1	33,3
	Desvio Padrão	2,6	2,0	2,7	2,8

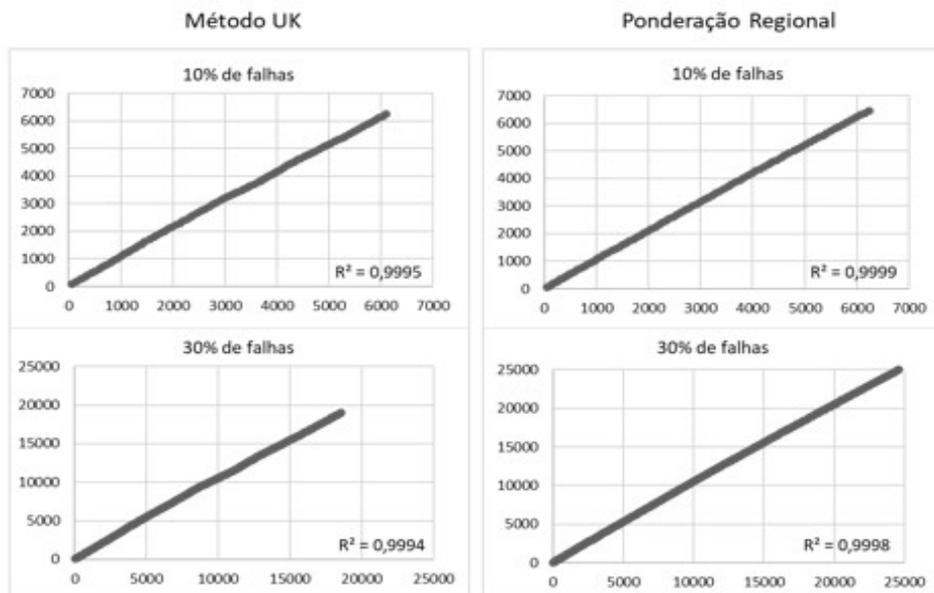
Para a temperatura mínima a diferença foi de até 0,2°C na ponderação regional e até 0,7°C na regressão múltipla (Tab. 6). Os valores de desvio padrão das amostras simuladas e observadas foram semelhantes para o método de ponderação regional, não excedendo a diferença de 0,1°C. Já para a regressão múltipla esse valor foi levemente maior (0,5°C).

**Tabela 6: Valores mínimos, máximos e desvio padrão dos dados de temperatura mínima simulados e observados**

		10%		30%	
		Simulado	Observado	Simulado	Observado
Ponderação Regional	Mínimo	8,9	8,7	8,8	8,7
	Máximo	21,7	21,8	21,5	21,7
	Desvio Padrão	3,4	3,4	3,4	3,3
Regressão Múltipla	Mínimo	9,0	8,7	8,5	8,7
	Máximo	21,1	21,8	21,9	21,7
	Desvio Padrão	3,9	3,4	3,8	3,3

A umidade relativa do ar, assim como a precipitação, somente apresentou boa correlação com o método de ponderação regional, 0,94 para 10% falhas e 0,92 para 30% de falhas. Embora o erro médio não tenha sido o menor observado nesse método, ele esteve abaixo de 1% de UR para ambos percentuais de falhas (-0,04 para 10% falhas e 0,34 para 30%). O método UK apresentou o segundo melhor coeficiente de determinação (0,79). Ainda que este seja menor que 80%, foi aplicado o teste de dupla massa a título de comparação com a ponderação regional.

O método de ponderação apresentou melhor ajuste da reta de dupla massa, sobretudo com relação aos dados com 30% de falhas, quando o método UK subestimou os valores simulados em relação aos observados (Fig. 7).



**Figura 7: Teste de dupla massa para os métodos de ponderação regional e ponderação regional com base em regressões lineares aplicado à umidade relativa do ar.**

Os valores máximos e mínimos simulados estiveram mais próximos dos observados no método da ponderação regional, assim como os valores de desvio padrão das amostras que não excederam a diferença de 4% (Tab. 7). Já para a regressão múltipla a diferença entre o desvio padrão da amostra simulada

e observada com 30% de falhas foi significativa, chegando a uma diferença de 3,1%.

**Tabela 7: Valores mínimos, máximos e desvio padrão dos dados de umidade relativa do ar simulados e observados**

		10%		30%	
		Simulado	Observado	Simulado	Observado
Ponderação Regional	Mínimo	40,0	41,7	39,4	42,3
	Máximo	94,0	92,2	87,1	86,7
	Desvio Padrão	11,9	11,5	11,5	11,5
Método UK	Mínimo	31,2	41,7	39,4	42,3
	Máximo	85,0	92,2	71,8	86,7
	Desvio Padrão	14,6	11,5	13,4	11,5

### Considerações finais

Nenhum dos testes aplicados para o preenchido de falhas da precipitação apresentaram coeficientes de determinação acima de 0,9. O resultado condiz com o observado por Acock e Pachepsky (2000) e com a característica espacial e temporal desse elemento climático que se distribui no espaço de maneira pontual e está exposto a variabilidades e oscilações naturais do clima que resultam em totais pluviométricos mensais distintos entre si ano a ano.

Os melhores coeficientes de determinação encontrados para a simulação da precipitação foram obtidos através dos métodos de ponderação regional e ponderação regional com base em regressões lineares. Ambos apresentaram consistência dos dados simulados satisfatória, embora a ponderação regional tenha apresentado resultados ligeiramente melhores, bem como valores extremos simulados mais próximos dos observados sendo este o método escolhido para o preenchimento das falhas de precipitação das estações meteorológicas presentes na bacia hidrográfica do Rio das Velhas e entorno.

Para a temperatura do ar (média, máxima e mínima), a ponderação regional também apresentou os melhores resultados, ainda que muito próximos daqueles gerados pela regressão múltipla. A simulação dessa variável por meio da ponderação e da regressão múltipla também foi satisfatória em outros estudos, em função do caráter contínuo de manifestação desta variável no espaço (YAGUCHI *et al.*, 2016; VENTURA *et al.*, 2016; BIA e FERRAZ, 2017; MELLO *et al.*, 2017). O teste de dupla massa para ambos atestou consistência dos dados. No entanto a ponderação regional apresentou resultados um pouco melhores, assim como valores extremos e de desvio padrão simulados mais próximos do observado, sendo também o método escolhido para o preenchimento de falhas da temperatura do ar.

A variável umidade relativa do ar apresentou coeficiente de determinação acima de 0,9 somente para o método de ponderação regional. O segundo maior valor desse coeficiente foi de 0,79 com o método UK. Embora ambos tenham sido comparados quanto ao teste de dupla massa, valores extremos simulados e desvio padrão, a ponderação regional apresentou melhores resultados

sendo o método escolhido para o preenchimento das falhas da umidade relativa do ar.

A princípio, seriam escolhidos os melhores métodos para cada variável individualmente, conforme recomenda Mello *et al.* (2017). No entanto, a ponderação regional foi o método que apresentou melhores resultados estatísticos para todas as variáveis abordadas sendo ele o escolhido para o preenchimento das falhas nas séries históricas das onze estações meteorológicas contidas na bacia hidrográfica do Rio das Velhas e entorno.

## Referências bibliográficas

ACOCK, M. C.; PACHEPSKY, Y. A. A. Estimating missing weather data for agricultural simulations using group method of data handling. *Journal Climate Applied Meteorology and CLimatology*, v. 39, p. 1176–1184, 2000.

AGÊNCIA NACIONAL DE ÁGUAS. *Análises recomendadas para a consistência de dados pluviométricos*. Brasília: ANA, Superintendência de Gestão da Rede Hidrometeorológica, 2011. 18 p.

ALLEN, R. J.; DeGAETANO, A. T. Estimating missing daily temperature extremes using an optimized regression approach. *International Journal of Climatology*, v. 21, p. 1305-1319, 2001.

BIER, A. A.; FERRAZ, S. E. T. Comparação de Metodologias de Preenchimento de Falhas em Dados Meteorológicos para Estações no Sul do Brasil. *Revista Brasileira de Meteorologia*, v. 32, n. 2, p. 215-226, 2017.

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, p. 324-339, 2018.

FANTE, K. P.; SANT'ANNA NETO, J. L. Técnicas estatísticas para a homogeneização de dados de temperatura e séries temporais climatológicas. *Revista Brasileira de Climatologia*, v. 18, p. 143-156, 2016.

KEMP, W. P. D.; BURNELL, D. G.; EVERSON, D. O.; THOMSON, A. J. Estimating missing daily maximum and minimum temperatures. *Journal Climate Applied Meteorology and CLimatology*, v. 22, p. 1587–1593, 1983.

MELLO, C. R.; SILVA, A. M. *Hidrologia: princípios e aplicações em sistemas agrícolas*. Lavras: Ed. UFLA, 2013. 455 p.

MELLO, Y. R.; KOHLS, W.; OLIVEIRA, T. M. N. Uso de diferentes métodos para o preenchimento de falhas em estações pluviométricas. *Boletim de Geografia*, Maringá, v. 35, p. 112-120, 2017.

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, p. 1186-1192, 2010.

PINTO, H. S.; ALFONSI, R. R. Estimativa das temperaturas médias, máximas e mínimas mensais no estado do Paraná, em função de altitude e latitude. In: *Caderno de ciências da terra*. Universidade de São Paulo. Instituto de Geografia. São Paulo/SP. Editora gráfica Cairu. V.52, p. 1–28, 1974.

TABONY, R.C. The estimation of missing climatological data. *Journal of Climatology*, v. 3, n. 3, p. 297-314, 1983.

TEEGAVARAPU, R. S.V.; CHANDRAMOULI, V. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, v. 312, p. 191-206, 2005.

TUNG, Y.K. Point rainfall estimation for a mountainous region. *Journal of Hydraulic Engineering*, v. 109, p. 1386-1393, 1983.

VENTURA, T. M.; OLIVEIRA, A. G.; MARQUES, H. O.; OLIVEIRA, R. S.; MARTINS, C. A.; FIQUEIREDO, J. M.; BONFANTE, A. G. Uma abordagem computacional para preenchimento de falhas em dados micro meteorológicos. *Revista Brasileira de Ciências Ambientais*, n. 27, p. 61-70, 2013.

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, p. 168-177, 2016.

VIEIRA, S.; 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, p. 189-204, 2018.

YAGUCHI, S. M.; MASSIGNAM, A. M.; RICCE, W. S.; PANDOLFO, C. Preenchimento de Falhas dos Dados Diários de Temperatura Máxima e Mínima do Ar. *Ciência e Natura*, Santa Maria, v. 38, p. 1419-1425, 2016.