

Artigo

## Comparação de Metodologias de Preenchimento de Falhas em Dados Meteorológicos para Estações no Sul do Brasil

Anderson Augusto Bier<sup>1</sup>, Simone Erotildes Teleginski Ferraz<sup>2</sup><sup>1</sup>*Programa de Pós-Graduação em Meteorologia, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil.*<sup>2</sup>*Departamento de Física, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil.*

Recebido em 25 de Novembro de 2015 - Aceito em 6 de Dezembro de 2016

### Resumo

A rede de estações meteorológicas no Brasil é muito recente, com poucas estações no país com mais de 100 anos de dados. Além disso, muitas das séries oriundas destas estações apresentam falhas, seja por falta de medição (no caso das estações convencionais), seja por falha nos equipamentos (no caso das estações automáticas). Encontram-se diferentes métodos na literatura para o preenchimento destas falhas. Neste trabalho são comparados seis métodos, os quais foram aplicados a séries de dados mensais conhecidas de temperatura média compensada e precipitação para estações meteorológicas localizadas no estado do Rio Grande do Sul, região sul do Brasil. Os resultados para temperatura média compensada indicaram os métodos de regressão linear múltipla e ponderação regional como os mais adequados para estimativas de dados faltantes enquanto que para precipitação não houve um método a se sobressair.

**Palavras-chave:** estações meteorológicas, dados faltantes, métodos de preenchimento de falhas, temperatura média compensada, precipitação.

## Comparison of Gap Filling Methodologies for Meteorological Data in Southern Brazil Stations

### Abstract

The network of weather stations in Brazil is very recent, with few stations in the country possessing data over more than 100 years. In addition, many of the series from these stations present missing data, either due to lack of measurement (in the case of conventional stations) or due to equipment failures (in the case of automatic stations). There are different methods in the literature for filling these gaps. In this work, six methods are compared and applied to series of known monthly data for mean compensated temperature and precipitation for weather stations located in the State of Rio Grande do Sul, in Southern Brazil. The results for mean compensated temperature suggest that multiple linear regression methods and regional weighing are best suited for estimating missing data, while for precipitation there was no superior method.

**Keywords:** meteorological stations, missing data, gap filling methodologies, compensated mean temperature, precipitation.

### 1. Introdução

Um dos grandes problemas que surgem ao se trabalhar com dados medidos em estações meteorológicas é a incompletude destes. Esta falta de informações pode ser devido a uma falha no equipamento (se este for automático) ou a uma falta de observador (se esta for uma estação convencional). Como estes dados serão utilizados para diversos fins de pesquisa, a necessidade de métodos para

preencher estas falhas é de grande interesse. Há uma ampla variedade de métodos para o preenchimento destas lacunas, desde técnicas simples como o cálculo da média aritmética entre dados oriundos de estações vizinhas, até métodos mais apurados, como redes neurais, as quais são caracterizadas por algoritmos computacionais robustos. De forma geral, os métodos são baseados em fórmulas estatísticas ou

em algoritmos elaborados com o intuito de gerar estimativas para os dados faltantes.

Atualmente com a necessidade de se estimar diversas variáveis meteorológicas, uma gama de métodos de preenchimento de falhas tem sido testada. Moffat *et al.* (2007), analisou 15 técnicas diferentes de preenchimento de dados de NEE (Net Ecosystem CO<sub>2</sub> Exchange). Eles mostraram que técnicas de regressão não-linear, look-up table, amostragem de distribuição marginal e o modelo semiparamétrico geralmente demonstram bom desempenho. A rede neural artificial (RNA) foi ligeiramente superior às outras técnicas, enquanto que a técnica de interpolação simples mostrou um desempenho moderado, mas consistente. Várias técnicas sofisticadas, como o filtro de Kalman, o método de imputação múltipla (MIM), entre outras, apresentaram baixa confiabilidade para valores anuais, indicando que o desenvolvimento adicional poderá ser necessário.

A sociedade americana de engenheiros civis (ASCE) sugere que, embora o desempenho de uma RNA para estimar dados ausentes de precipitação já tenha sido verificado, uma solução alternativa deve ser sugerida para os casos em que os dados disponíveis são insuficientes devido à dependência das RNAs de dados de alta qualidade e quantidade (Lee e Kang, 2015). Neste mesmo estudo os autores propõem cinco diferentes funções de Kernel para estimar dados faltantes, com resultados de alta qualidade provindos de interpolação. Aly *et al.* (2009), avaliou diferentes interpolações utilizando métodos determinísticos e estocásticos para preencher dados diários de precipitação. Xia *et al.* (1999) utilizaram seis métodos para interpolar dados de temperatura, pressão de vapor, vento e precipitação. Eles utilizaram métodos simples como a média aritmética simples, o método do inverso da distância, o método da razão normal e o melhor estimador simples, assim como métodos mais robustos como regressão linear múltipla e mínimo desvio absoluto. Eles concluíram que, para a precipitação média, os seis métodos apresentam resultados similares. O método do inverso da distância ponderada também é testado em Teegavarapu e Chandramouli (2005), onde, além deste, os autores utilizam outros três métodos para estimar precipitação, sendo estes o método ponderado do coeficiente de correlação, a RNA e o método estimador de Kriging, onde estes três últimos apresentaram os melhores resultados. Beckers e Rixen (2003) sugerem o uso de diferentes métodos multivariados (SVD, EOF, POD, etc.) para o preenchimento de falhas em dados meteorológicos e oceânicos. Os autores sugerem que estes métodos são mais simples por não necessitarem de nenhuma calibração prévia.

Este artigo discute a aplicabilidade de seis métodos de preenchimento de falhas em dados para estações meteorológicas do estado brasileiro do Rio Grande do Sul. Os métodos utilizados são os seguintes: regressão linear múltipla (Tabony, 1983), ponderação regional (Bertonni e Tucci, 2007; Alves *et al.*, 2006 apud Oliveira *et al.*, 2010; Paulhus e Kohler, 1952), interpolação do inverso da distância (Hub-

bard, 1994), método da razão normal (Paulhus e Kohler, 1952; Young, 1992), método tradicional do Reino Unido (Tabony, 1983) e média aritmética simples. Eles foram testados em séries mensais de temperatura média compensada e precipitação, porém a aplicabilidade deles pode se estender a outros parâmetros meteorológicos. Para cada um dos parâmetros meteorológicos utilizados (temperatura média compensada e precipitação) é retirado um ano de dados mensais de estações meteorológicas testes e através dos métodos de preenchimento de falhas são criadas estimativas para substituir os dados retirados. Desta forma as estimativas podem ser avaliadas comparando-as com os dados originalmente registrados.

## 2. Metodologia

### 2.1. Dados

Os dados utilizados neste trabalho são oriundos das estações convencionais do Instituto Nacional de Meteorologia (INMET), disponibilizados no Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP), além de estações pertencentes à Agência Nacional de Águas (ANA) e a Companhia Estadual de Energia Elétrica (CEEE). Foram utilizadas séries mensais de temperatura média compensada (TMC) e precipitação, com o período utilizado sendo entre os anos de 2003 e 2012. A TMC é descrita pela seguinte fórmula:

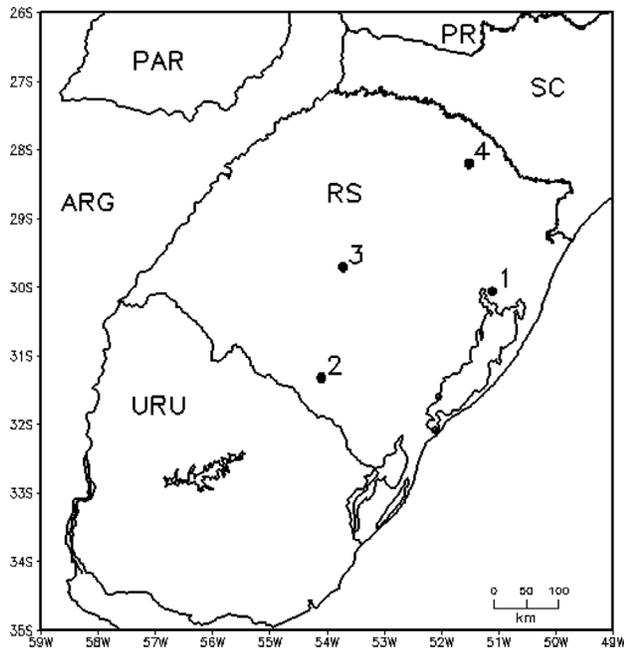
$$T_{MC} = (T_{max} + T_{min} + T_{12} + 2T_{24}) / 5 \quad (1)$$

sendo,  $T_{MC}$  a temperatura média compensada,  $T_{min}$  é a temperatura mínima do dia,  $T_{max}$  a temperatura máxima do dia,  $T_{12}$  a temperatura às 12 UTC e  $T_{24}$  a temperatura às 0 UTC.

Para podermos analisar os métodos de preenchimento de falhas de dados, se fazem necessárias estações meteorológicas testes. Para cada uma destas estações é retirado o período de um ano de dados mensais da série original e posteriormente são realizadas estimativas pelos diferentes métodos para substituir a série ao longo deste período. Os métodos, por sua vez, ao realizarem os cálculos das estimativas, necessitam de dados de estações vizinhas à estação teste. Desta forma, para ambas as variáveis, foram selecionadas como estações testes as seguintes estações meteorológicas convencionais do INMET: Porto Alegre, Bagé, Santa Maria e Lagoa Vermelha, todas localizadas no estado brasileiro do Rio Grande do Sul. Um mapa com a localização destas estações pode ser visualizado na Fig. 1.

Conforme Tabony (1983), alguns fatores importantes não de ser levados em conta na escolha das estações vizinhas:

- Cada parâmetro meteorológico deve ser estimado separadamente. Ou seja, as estações vizinhas usadas para um parâmetro (precipitação), não necessariamente serão as melhores a serem usadas para outro parâmetro (temperatura).



**Figura 1** - Localização das estações testes sobre o estado brasileiro do Rio Grande do Sul: 1) Porto Alegre, 2) Bagé, 3) Santa Maria e 4) Lagoa Vermelha. Sigla de estados: RS (Rio Grande do Sul), SC (Santa Catarina), PR (Paraná). Sigla de países: PAR (Paraguai), ARG (Argentina), URU (Uruguai).

- As estações vizinhas devem ser selecionadas de acordo com a correlação total ( $r$ ) com a estação teste. Observa-se que maiores correlações entre as estações vizinhas e teste tendem a gerar melhores estimativas.
- Uma dependência direcional pode ser imposta, isto é, se, por exemplo, doze estações vizinhas forem escolhidas que ao menos duas estejam em cada quadrante. Mas há de se levar em conta que nem sempre este esquema pode ser aplicado (por exemplo, próximo a regiões litorâneas).
- No cálculo das estimativas são levados em conta apenas os meses em que há sobreposição entre os dados da estação teste e estações vizinhas.

Considerando-se estes fatores, em um primeiro momento a seleção das estações vizinhas para cada variável meteorológica foi feita da seguinte maneira:

- TMC: foram selecionadas as doze estações vizinhas com maior correlação com a estação teste em um raio de 400 km partindo-se da estação teste.
- Precipitação: foram selecionadas as doze estações vizinhas com maior correlação com a estação teste em um raio de 100 km partindo-se da estação teste.

A diferença no tamanho dos raios escolhidos entre as duas variáveis se dá, principalmente, devido à precipitação ser um parâmetro mais variável espacial e temporalmente do que a temperatura, possuindo assim uma perda muito maior na correlação entre a estação teste e a estação vizinha com a distância, e também por serem estas as distâncias mínimas para que todas as estações testes possuíssem o mínimo de doze estações vizinhas na região de estudo. Na

Tabela 1 são listadas todas as estações (testes e vizinhas) utilizadas para a variável de TMC, assim como órgão pertencente, coordenadas geográficas e altitude de cada uma. As estações vizinhas ordenadas de acordo com a correlação com a estação teste são listadas na Tabela 2. Nas Tabelas 3 e 4 usa-se o mesmo molde das Tabelas 1 e 2 respectivamente, só que para a variável de precipitação. Nas Figs. 2 e 3 são mostradas a localização e a correlação das estações vizinhas com suas estações testes para as variáveis de TMC e precipitação respectivamente, onde se observam valores típicos da ordem de 0,9 para a correlação entre estações testes e vizinhas para a variável de TMC, e da ordem de 0,8 para a variável de precipitação, mesmo com os dados de TMC sendo espacialmente mais espaçados com relação aos de precipitação. Na seção de resultados e discussões, os números de estações vizinhas utilizadas serão reavaliados com intuito de definir o número mais adequado na estimativa de cada variável.

Vários dos métodos de preenchimento de falhas utilizados necessitam de um período de alguns anos de dados para a realização de médias, correlações e regressões que são utilizadas nos cálculos das estimativas. Para que as estimativas sejam consistentes, existe a necessidade de que haja uma boa disponibilidade de dados neste período, tanto nas estações teste quanto nas estações vizinhas. Portanto o período de dados tem que ser escolhido de forma que não haja uma grande quantidade de dados faltantes. Tabony (1983) sugere a utilização de um período de pelo menos 10 anos e com ao menos 50% dos dados disponíveis ao longo desse período. Desta forma, foi selecionado o período de dez anos de dados entre os anos de 2003 e 2012, devido a todas as estações apresentarem mais de 85% de seus dados disponíveis tanto para TMC quanto para precipitação durante esse intervalo de tempo. Para a avaliação dos métodos, os dados mensais de TMC e precipitação do ano de 2004 das estações testes foram retirados e substituídos pelas estimativas geradas pelos diferentes métodos, que desta maneira podem ser avaliados através de erros estatísticos entre a estimativa e o dado original. Tal ano foi escolhido por apresentar todos os dados disponíveis nas estações testes, tanto para precipitação quanto para TMC, possibilitando assim a avaliação das estimativas a partir de erros estatísticos e também pelas estações vizinhas apresentarem boa disponibilidade de dados, os quais são utilizados nos cálculos das estimativas.

## 2.2. Métodos de preenchimento de falhas

### 2.2.1. Regressão linear múltipla (RLM)

Uma das formas mais simples de se estimar dados faltantes é através da combinação linear de observações de estações meteorológicas vizinhas. O peso associado a cada observação pode ser determinado utilizando correlações parciais ou totais. Em geral, o método de correlações parciais entre estações (no qual se elimina o efeito das estações vizinhas) é a base para os métodos de regressão linear

**Tabela 1** - Estações meteorológicas utilizadas para TMC. As estações listadas em negrito correspondem às estações testes.

Nº	Estação	Código (OMM)	Órgão	Lat. (°)	Lon. (°)	Alt. (m)
<b>1</b>	<b>Bagé-RS</b>	<b>83980</b>	<b>INMET</b>	<b>-31,33</b>	<b>-54,10</b>	<b>242</b>
2	Bento Gonçalves-RS	83941	INMET	-29,15	-51,51	640
3	Bom Jesus-RS	83919	INMET	-28,66	-50,43	1048
4	Campos Novos-SC	83887	INMET	-27,38	-51,20	947
5	Caxias do Sul-RS	83942	INMET	-29,16	-51,20	760
6	Chapecó-SC	83883	INMET	-27,11	-52,61	679
7	Cruz Alta-RS	83912	INMET	-28,63	-53,60	472
8	Encruzilhada do Sul-RS	83964	INMET	-30,53	-52,51	428
9	Lages-SC	83891	INMET	-27,81	-50,33	937
<b>10</b>	<b>Lagoa Vermelha-RS</b>	<b>83916</b>	<b>INMET</b>	<b>-28,21</b>	<b>-51,50</b>	<b>840</b>
11	Passo Fundo-RS	83914	INMET	-28,21	-52,40	684
12	Pelotas-RS	83985	INMET	-31,78	-52,41	13
<b>13</b>	<b>Porto Alegre-RS</b>	<b>83967</b>	<b>INMET</b>	<b>-30,05</b>	<b>-51,16</b>	<b>47</b>
14	Rio Grande-RS	83995	INMET	-32,03	-52,10	2
<b>15</b>	<b>Santa Maria-RS</b>	<b>83936</b>	<b>INMET</b>	<b>-29,70</b>	<b>-53,70</b>	<b>95</b>
16	Santa Vitória do Palmar-RS	83997	INMET	-33,51	-53,35	24
17	Santana do Livramento-RS	83953	INMET	-30,83	-55,60	328
18	São Luiz Gonzaga-RS	83907	INMET	-28,40	-55,01	245
19	Uruguaiana-RS	83927	INMET	-29,75	-57,08	62
20	Urussanga-SC	83923	INMET	-28,51	-49,31	48

**Tabela 2** - Estações vizinhas utilizadas por cada uma das estações testes para a variável de TMC com base na numeração exibida na Tabela 1. São utilizadas as seguintes abreviações para as estações testes: Porto Alegre (PA), Bagé (BG), Santa Maria (SM) e Lagoa Vermelha (LV).

	Estações vizinhas
PA	8, 12, 20, 15, 2, 14, 1, 5, 7, 9, 11 e 10
BG	17, 8, 15, 12, 13, 19, 7, 2, 18, 14, 16 e 5
SM	7, 18, 1, 13, 2, 8, 19, 11, 12, 17, 5 e 10
LV	11, 4, 2, 7, 3, 9, 5, 6, 15, 13, 20 e 8

múltipla. Já o método das correlações totais, utilizando uma grande quantidade de estações vizinhas é capaz de reduzir os erros aleatórios, mas não os sistemáticos. Tabony (1983) propõe um método de preenchimento de falhas utilizando regressão linear múltipla, adotando como variáveis explicativas os dados de estações vizinhas próximas da estação com falhas. Este método pressupõe os seguintes pontos:

- Para verificar se a correlação entre a estação vizinha e a estação teste é estatisticamente significativa é necessário que  $r$  tenha uma distribuição normal e deste modo, testes tradicionais como o teste  $t$  podem ser utilizados. Para garantir esta simetria,  $r$  é convertido usando a transformação  $z$  de Fisher (que apresenta uma distribuição normal). Este cálculo é realizado para cada mês em separado.
- As estações vizinhas escolhidas devem ser classificadas e sua importância deve ser pesada em relação ao seu

valor de  $z$  ( $r$  convertido usando a transformação descrita no primeiro item).

- O período de sobreposição dos dados entre as estações deve ser considerado, incluindo o erro padrão do coeficiente  $z$  nos critérios de seleção. Assim, estações são classificadas não de acordo com o coeficiente  $z$  diretamente, mas de acordo com seu limite de confiança de 95%.

### 2.2.2. Ponderação regional (PR)

Este método de preenchimento de falhas é descrito por Alves *et al.* (2006 apud Oliveira *et al.*, 2010), por Bertoni e Tucci (2007 apud Oliveira *et al.*, 2010) e por Paulhus e Kohler (1952) com o intuito de preencher dados faltantes de precipitação. A falha nos dados na estação meteorológica de interesse é estimada da seguinte forma:

$$D_x = \frac{1}{n} \sum_{i=1}^n \frac{M_x}{M_i} D_i \quad (2)$$

sendo  $D_x$  o dado mensal faltante a ser estimado para a estação teste,  $D_i$  o dado ocorrido na estação vizinha de ordem “ $i$ ” no mês de ocorrência da falha na estação teste,  $M_x$  o dado médio mensal da estação teste,  $M_i$  o dado médio mensal da estação vizinha de ordem “ $i$ ” e  $n$  o número de estações vizinhas utilizadas no cálculo.

**Tabela 3** - Estações meteorológicas utilizadas para precipitação. As estações listadas em negrito correspondem às estações testes. As abreviações são referentes aos municípios: Bagé-RS (BG), Cachoeira do Sul-RS (CS), Casca-RS (CC), Dom Pedrito-RS (DP), Gravataí-RS (GT), Ibiçá-RS (IB), Jaguari-RS (JG), Júlio de Castilhos (JC), Lagoa Vermelha-RS (LV), Lavras do Sul-RS (LS), Montenegro-RS (MT), Pinheiro Machado-RS (PM), São Gabriel-RS (SG), São Jerônimo (SJ), São Pedro do Sul-RS (SP), Tupanciretã-RS (TP) e Viamão-RS (VI)

Nº	Estação	Código (ANA)	Orgão	Lat. (°)	Lon.(°)	Alt. (m)
1	Aceguá-RS	3154007	ANA	-31,87	-54,16	270
2	Antônio Prado-RS	2851003	ANA	-28,85	-51,28	630
<b>3</b>	<b>Bagé-RS</b>	<b>3154001</b>	<b>INMET</b>	<b>-31,33</b>	<b>-54,10</b>	<b>242</b>
4	Barra do Ribeiro-RS	3051023	ANA	-30,30	-51,31	5
5	Barro Vermelho (CS)	3053021	ANA	-30,14	-53,16	100
6	Bom Retiro (SG)	2954033	ANA	-29,99	-54,49	180
7	Butiá-RS	3051031	ANA	-30,16	-51,93	60
8	Caçapava do Sul-RS	3053022	ANA	-30,52	-53,49	420
9	Candelária-RS	2952034	ANA	-29,67	-52,79	40
10	Capão do Cedro (LV)	2851050	ANA	-28,09	-51,54	805
11	Casca-RS	2851005	CEEE	-28,57	-51,97	600
12	Caxias do Sul-RS	2951008	INMET	-29,16	-51,20	760
13	Charqueadas-RS	2951067	ANA	-29,95	-51,63	21
14	Clemente Argolo (LV)	2751017	ANA	-28,01	-51,45	950
15	Dom Pedrito-RS	3054002	ANA	-30,98	-54,68	120
16	Esmeralda-RS	2851043	ANA	-28,06	-51,19	974
17	Estância do Espantoso (BG)	3154006	ANA	-31,53	-54,29	140
18	Furnas do Segredo (JG)	2954005	ANA	-29,36	-54,50	300
19	Glorinha (GT)	2950016	ANA	-29,88	-50,79	77
20	Ibaré (LS)	3054019	ANA	-30,77	-54,22	310
21	Jaguari-RS	2954007	ANA	-29,51	-54,68	100
<b>22</b>	<b>Lagoa Vermelha-RS</b>	<b>2851014</b>	<b>INMET</b>	<b>-28,21</b>	<b>-51,50</b>	<b>840</b>
23	Lavras do Sul-RS	3053007	CEEE	-30,82	-53,90	300
24	Lombas (VI)	3050008	ANA	-30,05	-50,70	130
25	Nossa Senhora de Fátima (LV)	2851052	ANA	-28,46	-51,54	263
26	Palmares do Sul-RS	3050002	ANA	-30,25	-50,51	3
27	Paraíso (BG)	3153003	CEEE	-31,23	-53,90	368
28	Passo das Pedras (IB)	2851020	CEEE	-28,26	-51,86	750
29	Passo Fundo-RS	2852020	INMET	-28,21	-52,40	684
30	Passo Migliaivaca (CC)	2851022	ANA	-28,62	-51,87	380
31	Passo São Lourenço (CS)	2953037	ANA	-30,01	-53,02	20
32	Pinheiro Machado-RS	3153008	ANA	-31,58	-53,38	440
33	Ponte Toropi II (SP)	2954032	ANA	-29,65	-54,47	40
<b>34</b>	<b>Porto Alegre-RS</b>	<b>3051011</b>	<b>INMET</b>	<b>-30,05</b>	<b>-51,16</b>	<b>47</b>
35	Porto Garibaldi (MT)	2951024	ANA	-29,81	-51,39	15
36	Quevedos (JC)	2954019	ANA	-29,35	-54,07	408
37	Quitéria (SJ)	3052011	ANA	-30,42	-52,07	300
38	Sananduva-RS	2751007	ANA	-27,98	-51,78	687
<b>39</b>	<b>Santa Maria-RS</b>	<b>2953017</b>	<b>INMET</b>	<b>-29,70</b>	<b>-53,70</b>	<b>95</b>
40	Santo Antônio da Patrulha-RS	2950060	ANA	-29,82	-50,52	80
41	São Bernardo (TP)	2854013	ANA	-28,91	-54,06	530
42	São Gabriel-RS	3054018	ANA	-30,36	-54,32	120
43	São Sepé-RS	3053020	ANA	-30,19	-53,55	60
44	São Vendelino-RS	2951027	ANA	-29,37	-51,37	140

**Tabela 3** - Cont.

Nº	Estação	Código (ANA)	Orgão	Lat. (°)	Lon.(°)	Alt. (m)
45	Sapucaia do Sul-RS	2951028	ANA	-29,82	-51,16	20
46	Serrilhada (DP)	3154008	ANA	-31,44	-54,71	230
47	Tapejara-RS	2852046	ANA	-28,06	-52,00	672
48	Torquato Severo (DP)	3154003	ANA	-31,03	-54,18	390
49	Torrinhas (PM)	3153017	ANA	-31,31	-53,50	420
50	Três Passos II (LS)	3053023	ANA	-30,97	-53,97	280
51	Três Vendas-RS (DP)	3155001	ANA	-31,28	-55,04	240
52	Trinta e Cinco (DC)	2851028	CEEE	-28,39	-51,85	650

**Tabela 4** - Estações vizinhas utilizadas por cada uma das estações testes para a variável de precipitação com base na numeração exibida na Tabela 3. São utilizadas as seguintes abreviações para as estações testes: Porto Alegre (PA), Bagé (BG), Santa Maria (SM) e Lagoa Vermelha (LV).

	Estações vizinhas
PA	13, 45, 24, 19, 26, 40, 4, 44, 37, 7, 35 e 12
BG	48, 17, 23, 46, 27, 15, 50, 20, 32, 51, 49 e 1
SM	33, 6, 21, 18, 42, 5, 31, 36, 9, 41, 8 e 43
LV	28, 38, 10, 16, 25, 47, 52, 11, 14, 29, 2 e 30

### 2.2.3. Interpolação do inverso da distância (IID)

Hubbard (1994) utiliza este seguinte método para o preenchimento de falhas em dados meteorológicos:

$$D_x = \frac{\sum_{i=1}^n (D_i / d_i)}{\sum_{i=1}^n (1/d_i)} \quad (3)$$

sendo  $D_x$  o dado mensal faltante a ser preenchido na estação teste,  $D_i$  o dado ocorrido na estação vizinha de ordem “i” no mês de ocorrência da falha na estação teste e  $d_i$  a distância entre a estação teste e a estação vizinha de ordem “i”.

### 2.2.4. Método da razão normal (MRN)

Este método foi originalmente proposto por Paulhus e Koller (1952), foi modificado por Young (1992) e trata-se de uma ponderação dos dados registrados em estações vizinhas realizada da seguinte maneira:

$$D_x = \frac{\sum_{i=1}^n D_i W_i}{\sum_{i=1}^n W_i} \quad (4)$$

sendo  $D_x$  o dado mensal faltante a ser preenchido na estação teste,  $D_i$  o dado ocorrido na estação vizinha de ordem “i” no mês de ocorrência da falha na estação teste e  $W_i$  o peso ponderativo atribuído a cada estação vizinha de ordem “i”, o qual é descrito pela seguinte fórmula:

$$W_i = r_i^2 \left( \frac{n_i - 2}{1 - r_i^2} \right) \quad (5)$$

sendo  $r_i$  a correlação entre a estação teste e a estação vizinha de ordem “i” e  $n_i$  o número de meses em que ocorreu sobreposição de dados entre a estação teste e a estação vizinha de ordem “i”, ou seja, é o tamanho da série de dados envolvida no cálculo do coeficiente de correlação.

### 2.2.5. Método tradicional do Reino Unido (UK)

Segundo Tabony (1983), o método de preenchimento de falhas tradicionalmente utilizado pelo U.K. Meteorological Office é baseado em comparações com uma única estação vizinha. Tal estação vizinha é selecionada a partir de sua correlação com a estação teste. Para temperatura (TMC no nosso caso), a estimativa pelo método é feita para cada mês assumindo-se um valor constante de diferença entre a estação vizinha e a estação teste. Este valor é a diferença entre o valor médio do mês observado na estação teste e na estação vizinha ao longo de um certo período de tempo. Por exemplo, se no mês de julho a média mensal de TMC na estação teste é 0,2 °C mais quente em comparação com a estação vizinha, a estimativa para um mês de julho onde se reportou uma falha na estação teste será o valor registrado na estação vizinha neste mesmo mês mais 0,2 °C. Já para precipitação é assumida uma razão constante. Por exemplo, se em Janeiro a média mensal da chuva na estação teste é de 80 mm e na estação vizinha é de 100 mm, será assumido que a estimativa para a precipitação para uma falha ocorrida no mês Janeiro será 20% menor do que o valor registrado na estação vizinha, e não 20 mm menor como seria o caso de um valor constante.

### 2.2.6. Média aritmética simples (MA)

A média aritmética é simplesmente:

$$D_x = \frac{1}{n} \sum_{i=1}^n D_i \quad (6)$$

sendo  $D_x$  o dado mensal faltante a ser preenchido na estação teste,  $D_i$  o dado ocorrido na estação vizinha de ordem “i” no mês de ocorrência da falha na estação teste e  $n$  o número de estações vizinhas utilizadas no cálculo.

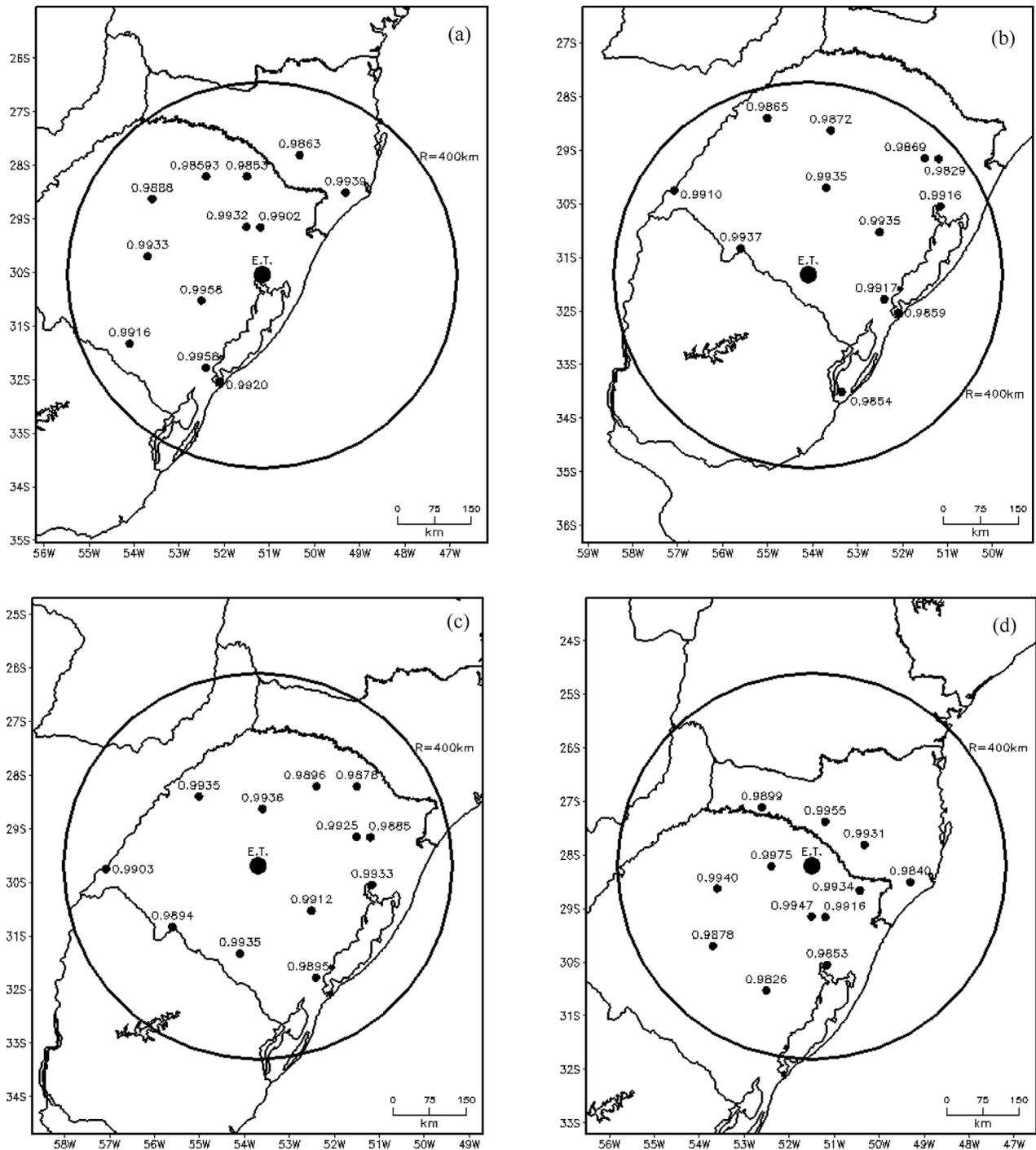


Figura 2 - Localização e valores de correlação das estações vizinhas com relação à estação teste (E.T.) para a variável de TMC. Estações testes: a) Porto Alegre, b) Bagé, c) Santa Maria e d) Lagoa Vermelha.

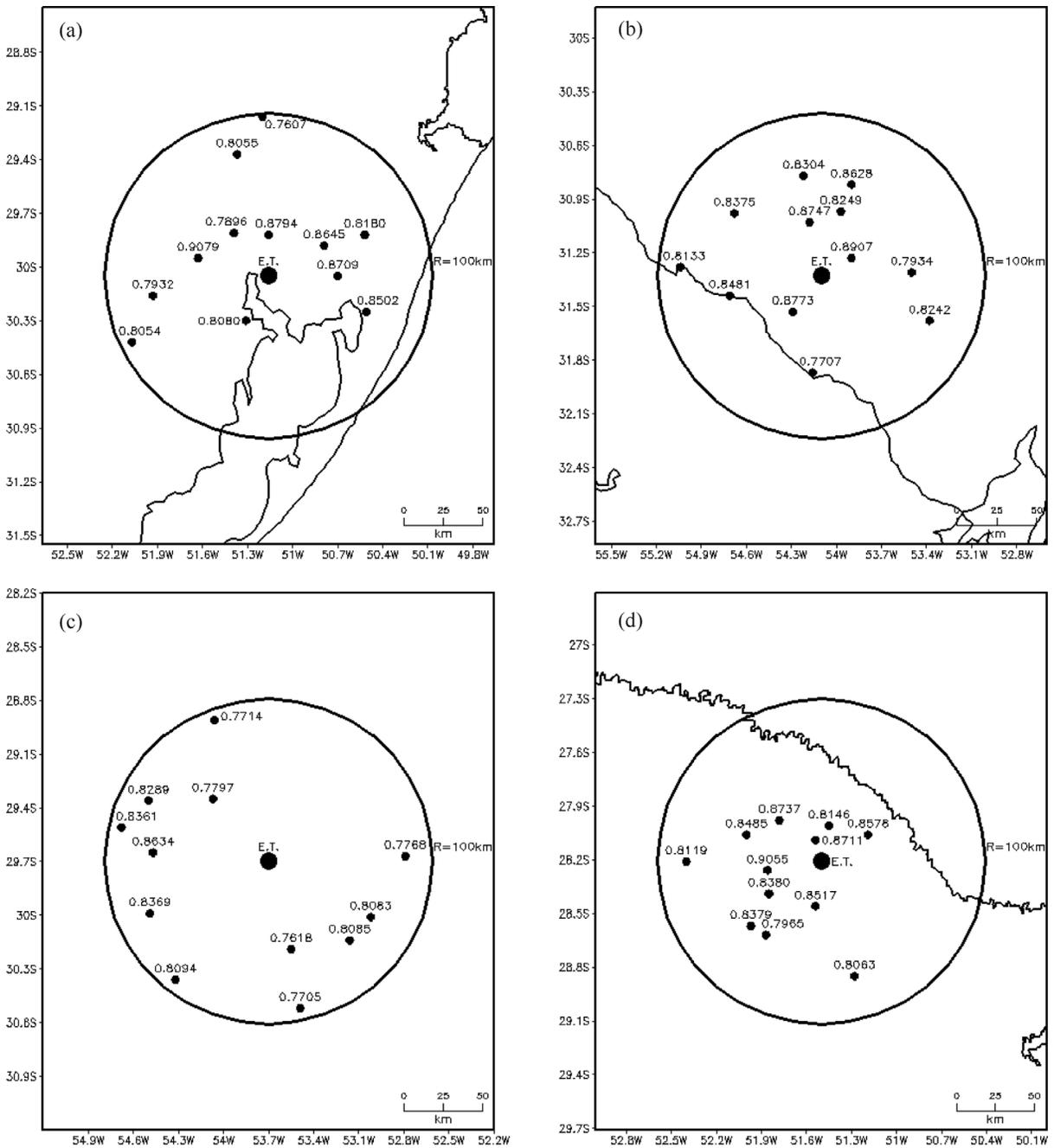
### 2.3. Erros estatísticos

Para a avaliação do desempenho dos diferentes métodos de preenchimento de falhas na efetuação das estimativas, foi realizada a medida do viés e do viés percentual, que quantificam subestimções e superestimções das estimativas com relação às observações médias, e também foi

efetuado o cálculo do erro absoluto médio e da raiz do erro quadrático médio que são medidas de precisão das estimativas. Conforme Déqué (2003), as fórmulas são descritas da seguinte forma:

- Viés ( $V$ )

$$V = \frac{1}{n} \sum_{i=1}^n (V_{est_i} - V_{obs_i}) \tag{7}$$



**Figura 3** - Localização e valores de correlação das estações vizinhas com relação à estação teste (E.T.) para a variável de precipitação. Estações testes: a) Porto Alegre, b) Bagé, c) Santa Maria e d) Lagoa Vermelha.

- Viés Percentual ( $VP$ )

$$VP = \frac{\sum_{i=1}^n (V_{est_i} - V_{obs_i})}{\sum_{i=1}^n V_{obs_i}} \quad (8)$$

- Erro absoluto médio ( $EAM$ )

$$EAM = \frac{1}{n} \sum_{i=1}^n |V_{est_i} - V_{obs_i}| \quad (9)$$

- Raíz do erro quadrático médio ( $REQM$ )

$$REQM = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_{est_i} - V_{obs_i})^2} \quad (10)$$

sendo  $n$  o número de meses observados, que neste trabalho corresponde aos 12 meses entre Janeiro de 2004 e Dezembro de 2004, nos quais os dados foram retirados da série original das estações testes.  $V_{obs_i}$  é o valor observado e  $V_{est_i}$  é o valor da estimativa em cada um dos meses.

### 3 Resultados e Discussões

Para facilitar a discussão, os nomes dos métodos de preenchimento de falhas são abreviados da seguinte forma: regressão linear múltipla (RLM), ponderação regional (PR), interpolação do inverso da distância (IID), método da razão normal (MRN), método tradicional do Reino Unido (UK) e média aritmética simples (MA).

A fim de verificar o número de estações vizinhas a gerar as melhores estimativas em média, o número inicial de 12 estações vizinhas foi variado entre 3 e 12 no cálculo das estimativas dos diferentes métodos em cada uma das estações testes, sempre com as estações vizinhas sendo selecionadas de acordo com a correlação com a estação teste, ou seja, ao utilizar-se 3 estações vizinhas nas estimativas, estas serão as 3 com maior correlação com sua estação teste, ou ao utilizar-se 4, estas serão as 4 melhores correlacionadas, e assim por conseguinte. Todas as metodologias apresentadas neste trabalho, com exceção do método UK (que necessita de apenas uma estação vizinha no seu cálculo), foram utilizadas nessa etapa de estimativas. As médias dos erros absolutos médios (*EAM*) ocorridos em todos estes métodos para o ano de 2004 para cada uma das estações testes, variando-se o número de estações vizinhas utilizadas, podem ser observadas na Tabela 5a para TMC e na Tabela 5b para precipitação. O número de estações vizinhas a gerarem as melhores estimativas varia um pouco de estação teste para estação teste conforme pode ser observado nas tabelas, assim como varia de método para método também (não mostrado). Porém, de maneira a facilitar a comparação entre os métodos de preenchimento de falhas, que será mostrada na seção a seguir, utilizamos o número de estações vizinhas a gerarem as melhores estimativas em média. Conforme podem ser observadas na Tabela 5a, para TMC, em média, as melhores estimativas foram geradas quando o número de estações vizinhas era igual a 4, enquanto que para precipitação esse número foi de 7 estações vizinhas. Desta maneira, tais números de estações vizinhas foram utilizados para a comparação entre os diferentes métodos de preenchimento de falhas para as variáveis de TMC e precipitação respectivamente. Quanto ao método UK, no qual se utiliza apenas uma estação vizinha, a estação vizinha selecionada para a efetuação da estimativa foi a de maior correlação com a estação teste.

#### 3.1. TMC

Na Tabela 6 podem ser verificados os erros absolutos médios (*EAM*) e as raízes do erro quadrático médio (*REQM*), gerados por cada uma das estimativas dos métodos de preenchimentos de falhas com relação aos valores originais de cada uma das estações testes e na média entre todas elas. Em praticamente todas as estações testes observa-se um padrão com os métodos PR, RLM e UK produzindo os menores valores em ambos os erros estatísticos, portanto, gerando as estimativas mais precisas. Os erros absolutos médios destes métodos em média ficaram entre

**Tabela 5** - Média dos erros absolutos médios das estimativas dos métodos RLM, PR, IID, MRN e MA para o ano de 2004 para a) TMC (em °C) e para b) precipitação (em mm) para as estações de Porto Alegre (PA), Bagé (BG), Santa Maria (SM), Lagoa Vermelha (LV) e média para as quatro estações testes (M), variando-se o número de estações vizinhas utilizadas nos cálculos (n).

(a)					
n	PA	BG	SM	LV	M
3	0,813	0,481	0,399	0,253	0,4865
4	0,693	0,454	0,268	0,397	<b>0,4530</b>
5	0,813	0,599	0,407	0,196	0,5038
6	0,801	0,642	0,504	0,186	0,5333
7	0,847	0,646	0,460	0,179	0,5330
8	0,944	0,606	0,521	0,207	0,5695
9	0,913	0,692	0,550	0,272	0,6068
10	0,994	0,692	0,579	0,354	0,6548
11	0,999	0,625	0,651	0,414	0,6723
12	1,042	0,571	0,693	0,413	0,6798

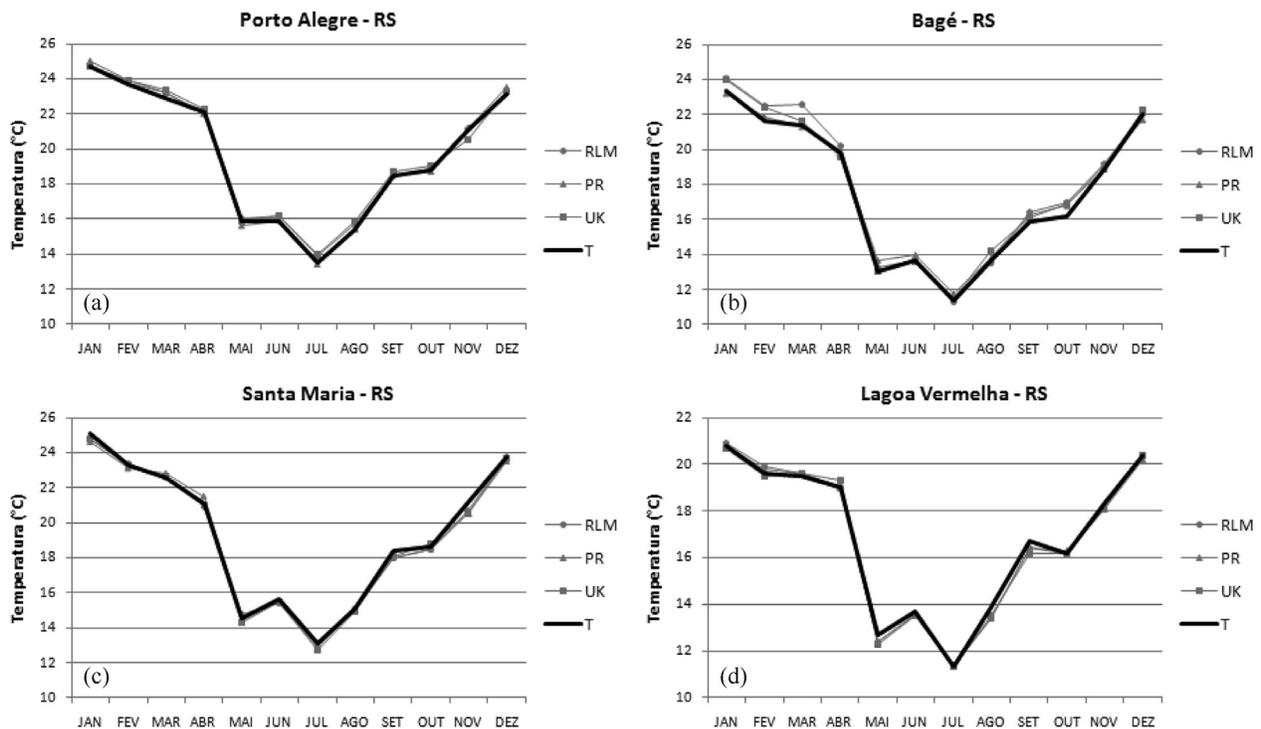
(b)					
n	PA	BG	SM	LV	M
3	16,94	20,09	35,95	17,18	22,538
4	16,34	20,50	32,79	16,48	21,527
5	14,36	21,54	29,61	16,08	20,349
6	13,58	21,83	28,73	17,21	20,337
7	12,07	21,66	28,02	18,50	<b>20,063</b>
8	12,80	21,99	27,38	18,67	20,208
9	13,85	22,53	27,65	19,37	20,850
10	14,98	22,60	27,33	20,05	21,239
11	15,32	22,75	27,49	20,29	21,463
12	14,96	23,04	26,95	20,87	21,454

0,221 °C e 0,252 °C enquanto os valores da *REQM* ficaram entre 0,277 °C e 0,322 °C em média. Tais estimativas contrastam com as dos métodos IID, MRN e MA, os quais tiveram valores de erros estatísticos aproximadamente três vezes superiores em média.

Nas Figs. 4 a, b, c e d são apresentados os gráficos das estimativas para TMC, geradas para cada um dos meses do ano de 2004 pelos três melhores métodos estimadores de cada uma das estações testes: Porto Alegre, Bagé, Santa Maria e Lagoa Vermelha respectivamente. Como já observado na Tabela 6, os métodos RLM, PR e UK possuem as melhores estimativas de forma geral. Pode se adquirir uma noção quantitativa da eficácia destes métodos, observando que o erro absoluto de estimativa máximo, nestes casos, é de 0,9 °C para a estação meteorológica teste de Bagé no mês de fevereiro, sendo que de forma geral, os erros absolutos de estimativa são bem mais próximos de 0 °C do que deste valor. As estimativas para as quatro estações meteo-

**Tabela 6** - Erro absoluto médio (*EAM*) e raiz do erro quadrático médio (*REQM*) das estimativas para cada um dos métodos em cada uma das estações testes para a variável de TMC. As melhores estimativas estão indicadas em negrito.

Método	Porto Alegre		Bagé		Santa Maria		Lagoa Vermelha		Média	
	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>
RLM	0,158	0,202	0,450	0,579	0,192	0,247	0,175	0,225	0,244	0,313
PR	0,158	0,206	0,300	0,358	0,258	0,325	0,167	0,220	<b>0,221</b>	<b>0,277</b>
IID	1,250	1,260	0,533	0,582	0,375	0,454	0,525	0,580	0,671	0,719
MRN	1,242	1,251	0,575	0,618	0,292	0,375	0,642	0,702	0,688	0,737
UK	0,283	0,342	0,308	0,405	0,225	0,287	0,192	0,253	0,252	0,322
MA	1,067	1,080	0,558	0,609	0,267	0,351	0,683	0,739	0,644	0,695



**Figura 4** - Gráficos com as estimativas para as séries mensais de TMC do ano de 2004 com os três melhores métodos para cada uma das estações testes: a) Porto Alegre, b) Bagé, c) Santa Maria e d) Lagoa Vermelha.

rológicas testes são bem pouco enviesadas. Somente a estação de Bagé apresenta valores um pouco mais elevados de viés, em média superestimados, com valores entre 0,2 °C e 0,4 °C nos métodos em questão (RLM, PR e UK), o que é possível constatar na análise do gráfico também (Fig. 4b).

Outros trabalhos encontrados na literatura já observaram os métodos de RLM e PR como sendo bons estimadores para temperatura, como é o caso de Fante e Neto (2013), onde foram encontrados valores inferiores à ordem de 0,9 °C para estimativas geradas pelo método PR, e para diversos trabalhos como Tabony (1983), Kemp *et al.* (1983), Eischeid *et al.* (1995), Xia *et al.* (1999) e Kashani *et al.* (2012), nos quais, métodos baseados em regressões lineares múltiplas obtiveram bons resultados.

### 3.2. Precipitação

Na Tabela 7 são verificados os erros absolutos médios (*EAM*) e as raízes do erro médio quadrático (*REQM*) das estimativas de cada método para precipitação em cada uma das estações testes. Na média entre todas as estações, o método que apresentou os melhores resultados foi o PR, com valores médios de *EAM* de 17,192 mm e de *REQM* de 22,181 mm, com os métodos IID, MA e MRN também obtendo resultados bastante próximos a estes nas medidas dos erros estatísticos. Considerando-se a precipitação mensal média de 100,65 mm para as quatro estações testes em questão no ano de 2004, os valores de erros absolutos médios das estimativas representam erros absolutos percentuais médios de 17,1% para o método PR, 17,9% para o IID, de 18,1% para o MA e 18,6% para o MRN.

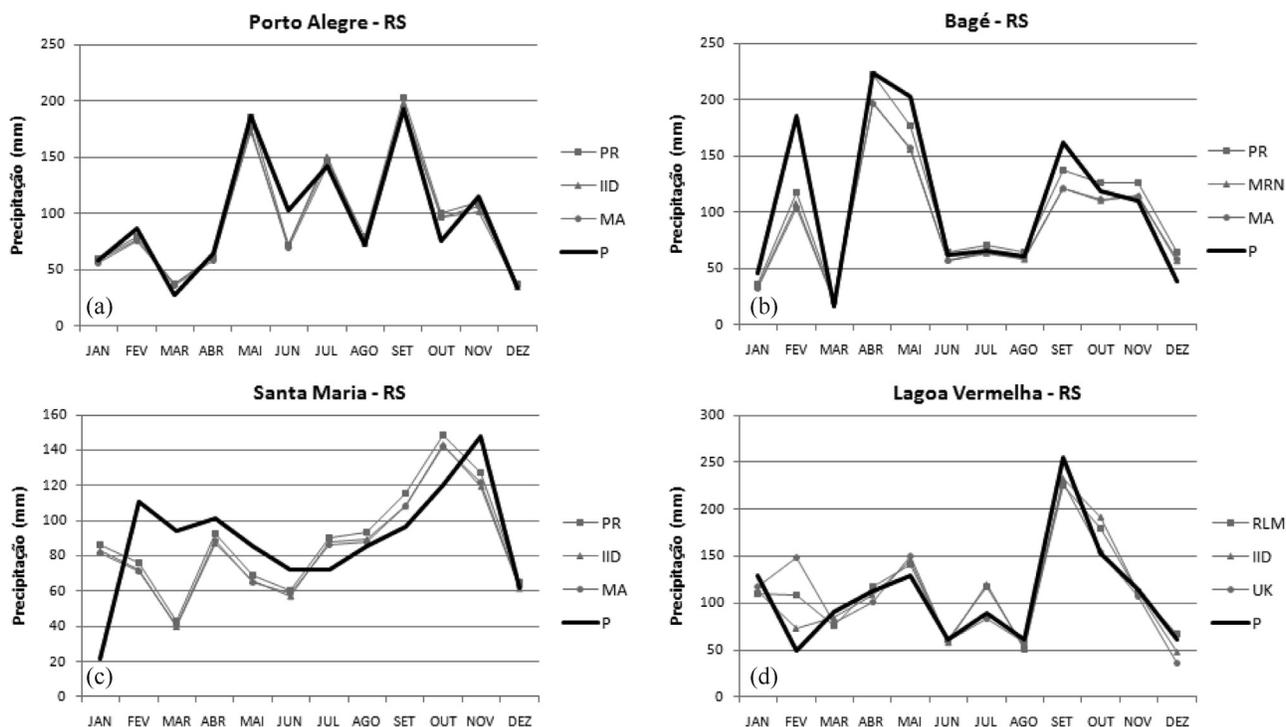
**Tabela 7** - Erro absoluto médio (*EAM*) e raiz do erro quadrático médio (*REQM*) das estimativas para cada um dos métodos em cada uma das estações testes para a variável de precipitação. As melhores estimativas estão indicadas em negrito.

Método	Porto Alegre		Bagé		Santa Maria		Lagoa Vermelha		Média	
	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>	<i>EAM</i>	<i>REQM</i>
RLM	17,32	22,41	22,42	27,13	26,73	44,48	17,96	23,95	21,106	29,493
PR	9,03	12,48	16,37	24,43	23,74	29,59	19,63	22,23	<b>17,192</b>	<b>22,181</b>
IID	9,88	13,08	22,43	33,54	23,83	29,85	15,73	19,32	17,967	23,947
MRN	10,38	13,22	21,00	30,82	24,18	30,85	19,27	22,43	18,708	24,331
UK	16,77	22,92	26,56	32,04	46,52	54,60	18,72	31,75	27,140	35,329
MA	9,04	12,82	21,20	31,37	23,13	29,31	19,69	22,63	18,267	24,033

Nas Figs. 5 a, b, c e d vemos os gráficos das estimativas para precipitação, geradas pelos três melhores métodos para cada uma das estações testes em cada um dos meses do ano de 2004. Observa-se que as estimativas são visualmente menos eficientes em comparação com as séries de TMC, muito em função de os dados de precipitação entre diferentes estações vizinhas serem menos correlacionados entre si em comparação aos dados de TMC, o que gera estimativas menos relacionadas com a série original e com erros de estimativas relativamente maiores. Mesmo assim, tais estimativas ainda conseguem representar da maneira razoavelmente boa à variação mensal da precipitação originalmente registrada nas estações testes, o que pode ser constatado ao observarmos a repetição dos padrões de picos de precipitação mensal entre as séries originais e suas estimativas nos gráficos. Quanto ao enviesamento das esti-

mativas mostradas nas figuras (as 3 melhores de cada estação teste) com relação aos valores originais, podemos constatar valores em média subestimados para as estações testes de Bagé (Fig. 5b), com valores de viés percentuais médios entre -15,24% a -4,8%, e de Santa Maria (Fig. 5c), com valores entre -5,21% e -0,04%. Por outro lado, para a estação teste de Lagoa Vermelha (Fig. 5d) os valores são, em média, levemente superestimados, com medidas de viés percentual entre 1,38% e 4,72%. Já para a estação teste de Porto Alegre (Fig. 5a) os valores de viés percentuais não foram muito significativos, ficando entre -3,01% e 1,16%.

Trabalhos anteriores, como Paulhus e Kohler (1952) e Oliveira *et al.* (2010), indicaram o método PR como sendo o mais adequado para a estimativa de precipitação. Tal método também apresentou as melhores estimativas em média neste trabalho, porém não obteve resultados signifi-



**Figura 5** - Gráficos com as estimativas para as séries mensais de precipitação do ano de 2004 com os três melhores métodos para cada uma das estações testes: a) Porto Alegre, b) Bagé, c) Santa Maria e d) Lagoa Vermelha.

cativamente melhores em comparação com os outros métodos apresentados. O método UK, por outro lado, apresentou erros de estimativa significativamente elevados para praticamente todas as estações testes, não se mostrando muito eficaz na estimativa de dados de precipitação.

#### 4. Conclusões

Com a realização de estimativas com diferentes métodos para estações meteorológicas localizadas no estado do Rio Grande do Sul, os métodos de preenchimento de falhas para dados meteorológicos que apresentaram melhores resultados para TMC em média foram o PR, o RLM e o UK com erros absolutos médios de estimativas próximos aos 0,2 °C e valores de raiz do erro médio quadrático entre 0,25 °C a 0,35 °C, mostrando significativa diferença em comparação aos outros três métodos, enquanto que para precipitação as estimativas geradas pelos diferentes métodos não apresentaram grandes diferenças entre si, com os métodos de PR, IID, MA e MRN apresentando uma pequena vantagem em comparação aos outros dois. Observaram-se estimativas relativamente melhores sendo geradas para a variável de TMC em comparação com a precipitação, devido às séries de TMC serem mais correlacionadas entre si, o que não faz com que as estimativas para precipitação percam sua validade.

Tais resultados são uma amostra de que é possível gerar boas estimativas para dados mensais a partir de métodos estatísticos simples para diferentes variáveis meteorológicas, principalmente em locais que possuam uma disponibilidade razoavelmente boa de estações meteorológicas, e desta forma, séries de dados com bons níveis de correlação entre si, como foi encontrado neste trabalho para o estado brasileiro do Rio Grande do Sul.

#### Referências

- ALVES, A.; KOBAYAMA, M.; SILVA, R.V.; CHECCHIA, T. Análise de dados hidrológicos na região do município de Alfredo Wagner/SC. In: Simpósio de Recursos Hídricos do Sul, 1, 2006, Curitiba. **Anais...** Porto Alegre: Associação Brasileira de Recursos Hídricos, 2006. CD-Rom.
- ALY, A.; PATHAK, C.; TEEGAVARAPU, R.S.V.; ALQUIST, J.; FUELBERG, H. Evaluation of improvised spatial interpolation methods for infilling missing precipitation records. In: **World Environment Water Resources Congress 2009: Great Rivers**. 2009. p. 1-10.
- BECKERS, J.; RIXEN, M. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets\*. **Journal of Atmospheric and Oceanic Technology**, v. 20, n. 12, p. 1839-1856, 2003.
- BERTONI, J.C.; TUCCI, C.E.M. Precipitação. In: TUCCI, C.E.M. **Hidrologia: Ciência e Aplicação**. Porto Alegre: UFRGS, 2007. p.177-241.
- DÉQUÉ, M. Continuous variables. In: JOLLIFE, I.T; STEPHENSON, D.B. **Forecast Verification: A Practitioner's Guide in Atmospheric Science**. Willey, 2003. p. 97-119.
- EISCHEID, J.K.; BAKER, C.B.; KARL, T.R.; DIAZ, H.F. The quality control of long-term climatological data using objective data analysis. **Journal of Applied Meteorology**, v. 34, n. 12, p. 2787-2795, 1995.
- FANTE, K.P.; NETO, J.L.S. Aplicação de técnicas geo-estatísticas em análises climatológicas: metodologias para preenchimento de falhas. In: Simpósio Brasileiro De Geografia Física Aplicada, 15, 20013, Vitória. **Anais...** Vitória: 2013. p. 782-790.
- HUBBARD, K.G. Spatial variability of daily weather variables in the high plains of the USA. **Agricultural and Forest Meteorology**, v. 68, n. 1, p. 29-41, 1994.
- KASHANI, M.H.; DINPASHOH, Y. Evaluation of efficiency of different estimation methods for missing climatological data. **Stochastic Environmental Research and Risk Assessment**, v. 26, n. 1, p. 59-71, 2012.
- KEMP, W.P.; BURNELL, D.G.; EVERSON, D.O.; THOMSON, A.J. Estimating missing daily maximum and minimum temperatures. **Journal of climate and applied meteorology**, v. 22, n. 9, p. 1587-1593, 1983.
- LEE, H.; KANG, K. Interpolation of Missing Precipitation Data Using Kernel Estimations for Hydrologic Modeling. **Advances in Meteorology**, v. 2015, 2015.
- 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 pluviométrica anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 11, p. 1186-1192, 2010.
- PAULHUS, J.L.H.; KOHLER, M.A. Interpolation of missing precipitation records. **Monthly Weather Review**, v. 80, n. 5, p. 129-133, 1952.
- 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, n. 1, p. 191-206, 2005.
- XIA, Y.; FABIAN, P.; STOHL, A.; WINTERHALTER, M. Forecast climatology: estimation of missing values for Bavaria, Germany. **Agricultural and Forest Meteorology**, v. 96, n. 1, p. 131-144, 1999.
- YOUNG, K.C. A three-way model for interpolating for monthly precipitation values. **Monthly Weather Review**, v. 120, n. 11, p. 2561-2569, 1992.