

# Análise de métodos para o preenchimento de falhas aplicados em séries de dados pluviométricos do estado do Paraná (Brasil)

## Analysis of imputing missing data methods applied to rainfall data series in Paraná state (Brazil)

Deise Fabiana Ely\*, Leila Limberger\*\*, Fabiana Bezerra Mangili\*\*\*, Paulo Gamero\*\*\*, Moacir Schmengler\*\*\*

Universidade Estadual de Londrina, e-mail: [deise.ely@gmail.com](mailto:deise.ely@gmail.com)

\*\*Universidade Estadual do Oeste do Paraná - UNIOESTE, e-mail: [leila.limberger@gmail.com](mailto:leila.limberger@gmail.com)

\*\*\* Parque Tecnológico de Itaipu, e-mailS: [fabianamangili@gmail.com](mailto:fabianamangili@gmail.com), [paulo.gamero@pti.org.br](mailto:paulo.gamero@pti.org.br), [moacir.schmengler@pti.org.br](mailto:moacir.schmengler@pti.org.br)

DOI: <http://dx.doi.org/10.5380/raega.v51i0.72944>

### Resumo

Em virtude de a Climatologia necessitar de uma série de dados consistente e sem falhas para análises climáticas e devido à grande quantidade de dados com falhas disponibilizados no Brasil, faz-se necessário avaliar métodos de preenchimento de falhas para viabilizar a utilização de séries longas de dados. O artigo objetiva avaliar qual método de preenchimento de falhas é mais adequado para estações meteorológicas localizadas no estado do Paraná. Para tanto, foram testados quatro métodos de preenchimento de falhas aplicados a 10 estações meteorológicas do Estado, integrantes de cinco regiões homogêneas de precipitação. Os métodos testados foram IDW, Inverso da Distância, Média e Ponderação Regional. Foram suprimidos os dados diários de anos classificados como habituais pelo método gráfico cronológico de tratamento da informação (MGCTI) e posteriormente preenchidos com os métodos sob análise. Posteriormente os valores preenchidos foram comparados com os valores removidos (registrados na estação). O método de preenchimento de falhas que demonstrou melhor associação com os dados registrados foi o IDW, seguido da Média. De forma geral os métodos de preenchimento de falhas testados subestimam os valores diários superiores a 20mm e superestimam os menores de 20mm, sendo que o IDW foi o que mais se aproximou dos valores acima de 40mm. Os métodos tiveram mais sucesso em preencher falhas nas estações da faixa sul do Estado, enquanto nas estações próximas ao litoral foram registradas as maiores diferenças entre os valores preenchidos e registrados.

**Palavras-chave:** Inverso da Distância; Inverso da Distância Ponderada (IDW); Ponderação Regional.

### Abstract

Due to the fact that Climatology needs a consistent and without missing data series for climate analysis and due to the large amount of missing data in Brazilian meteorological stations records, it is necessary to evaluate imputing missing data methods to make viable some long data series. This article aims to evaluate which imputing missing data method is most suitable for meteorological located in Paraná State. For this purpose, four imputing missing data methods were tested applied to 10 meteorological stations in the State,

members of five rainfall homogeneous regions. The tested methods were IDW, Distance Inverse, Average and Regional Weighting. The daily data of years classified as habitual by the Chronological Graphical Method of Treatment of Information (MGCTI) were suppressed and later filled with the methods under analysis. Subsequently, the completed values were compared with the removed values (registered at the station). The imputing missing data method that showed the best association with the recorded data set was the IDW, followed by the Average. In general, imputing missing data methods tested underestimate the daily values above 20mm and overestimate values below 20mm, and the IDW was the one that came closest to the values above 40mm. The methods were more successful in imputing data in stations in the South of the State, while the greatest differences between the filled and recorded values were recorded in the stations close to the coast.

**Keywords:** Inverse Distance; Inverse Distance Weighting (IDW); Regional Weighting.

## I. INTRODUÇÃO

A observação da atmosfera e a análise de eventos climáticos se concretizam por meio de dados atmosféricos (pressão do ar, umidade relativa do ar, temperatura do ar, precipitação, dentre outros). Dubreuil et al. (2012) esclarecem que a variabilidade de uma componente atmosférica pode se dar mais ou menos em torno de um valor de referência central (normal climatológica, por exemplo e por meio de sua análise é possível observar padrões e anomalias, identificando os anos mais ou menos quentes, ou mais úmidos e secos.

Para o estudo da variabilidade climática de um lugar é fundamental, portanto, uma base de dados amostral representativa, na qual os valores medidos (observados) devem ter um rigor para que possam representar o objeto estudado.

No Brasil existem diversos institutos que realizam a coleta dos dados atmosféricos, disponibilizando-os por meio de plataformas eletrônicas. Muitas vezes, porém, as séries históricas, além de apresentarem erros inerentes às medições, também registram falhas ou ausência de dados. Neste sentido, García-Peña; Arciniegas-Alarcón e Barbin (2014), ao abordarem o problema das falhas em séries históricas de dados, apontam que essas podem ocorrer pelo não registro dos dados devido à ausência, férias, à falta de capacitação dos operadores dos postos meteorológicos para o monitoramento das estações automáticas ou convencionais, problemas nos equipamentos de registro ou na transmissão dos dados de estações automáticas, além de outras adversidades.

Para sanar estas lacunas nas séries temporais vários métodos estatísticos são aplicados. Vieira et al. (2018) enfatizam que os métodos mais usuais para esta finalidade são: ponderação regional, regressão linear, regressão múltipla e potencial; além das redes neurais e de metodologias geoestatísticas como a krigagem. Dentre estes métodos, a ponderação regional e a regressão linear exigem estações próximas que disponham de dados para o estabelecimento da correlação com aqueles que se deseja preencher.

Vieira et al. (2018) ainda indicam que, quando o preenchimento de dados faltantes é realizado em uma série histórica, é importante que a inserção destas estimativas não altere as suas principais características, como variabilidade, movimentos sazonais e correlações com outras séries de dados que se deseja realizar associações.

Aieb et al. (2019) esclarecem que existem duas possibilidades para o preenchimento de falhas: 1) a técnica de eliminar o dado faltante, ou seja, excluir as células na matriz que identifica o dado faltante; indicada para situações em que o número de ausências for muito baixo e 2) a técnica da imputação de dados faltantes que consiste em substituir as lacunas por valores estimados por meio de vários métodos. Pode-se optar por um método de imputação simples ou múltiplo. A escolha por um método simples é indicada quando a porcentagem dos dados faltantes for menor que 5%, podendo-se substituir o valor ausente pela média de todas as observações medidas na mesma série. A regressão linear também é identificada como um método simples que estabelece a correlação entre dados observados de estações vizinhas e daquela com falhas; indicado quando se tem entre 5 e 10% de dados faltantes. A imputação múltipla é um tratamento estatístico que utiliza vários bancos de dados similares, fornecendo uma combinação de valores para a substituição do ausente.

Em virtude das características dos dados climáticos e, principalmente, a precipitação que possui uma distribuição aleatória, independente do método, o problema da confiabilidade dos resultados obtidos sempre permanece; não existindo uma técnica ideal de estimativa.

O presente artigo é parte de uma análise mais ampla das precipitações diárias no estado do Paraná, registradas no período de 1977 a 2018 e tem por objetivo testar diferentes métodos estatísticos nos processos de preenchimento de falhas de dados diários de precipitação de 10 postos pluviométricos. O método que demonstrar o melhor desempenho será utilizado para preencher as falhas identificadas nos dados diários de 271 postos pluviométricos distribuídos no Estado, que constituem a base para a análise de extremos climáticos.

## II. MATERIAIS E MÉTODOS

Segundo Limberger e Ely (2019), a precipitação no estado do Paraná é distribuída em cinco regiões homogêneas: Centro-Sul, Oeste/Sudoeste, Leste, Centro-Oeste e Norte; identificadas pelo agrupamento da análise de Cluster, baseada em um conjunto de dados de 330 estações pluviométricas para o período de 1977 a 2016. Para o teste dos métodos estatísticos para o preenchimento de falhas foram escolhidas duas estações meteorológicas representativas de cada região homogênea com dados pluviométricos diários completos (Quadro 1).

Quadro 1 – Regiões homogêneas de precipitação no estado do Paraná (LIMBERGER e ELY, 2019), anos habituais e postos meteorológicos adotados para análise do desempenho dos métodos estatísticos utilizados para o preenchimento de falhas.

Região	Anos Habituais	Postos
Centro-Sul	2002 e 2004	IAPAR – Palmas ANA – Inácio Martins/Inácio Martins
Oeste/Sudoeste	1994 e 2010	IAPAR – Francisco Beltrão ANA – São Miguel do Iguaçu/São Miguel do Iguaçu
Leste	1981, 1982 e 2013	IAPAR – Guaraqueçaba ANA – Piraquara/Piraquara
Centro-Oeste	1981 e 1995	IAPAR – Palotina ANA – Formosa do Oeste/Formosa do Oeste
Norte	1995 e 1996	IAPAR – Londrina ANA – Cidade Gaúcha/Cidade Gaúcha

A identificação dos anos-padrões (Quadro 1) foi realizada por meio do método gráfico cronológico de tratamento da informação (MGCTI) do tipo Matriz de Bertin (NOUACEUR; LAIGNEL; TURKI, 2014; ELY, 2019) e que propiciou a escolha dos anos com dados diários completos e sem anomalias positivas ou negativas de precipitação (Figura 1). Os dados pluviométricos diários destes anos foram suprimidos (4,76% da série histórica) e, posteriormente, preenchidos e comparados com os dados observados.

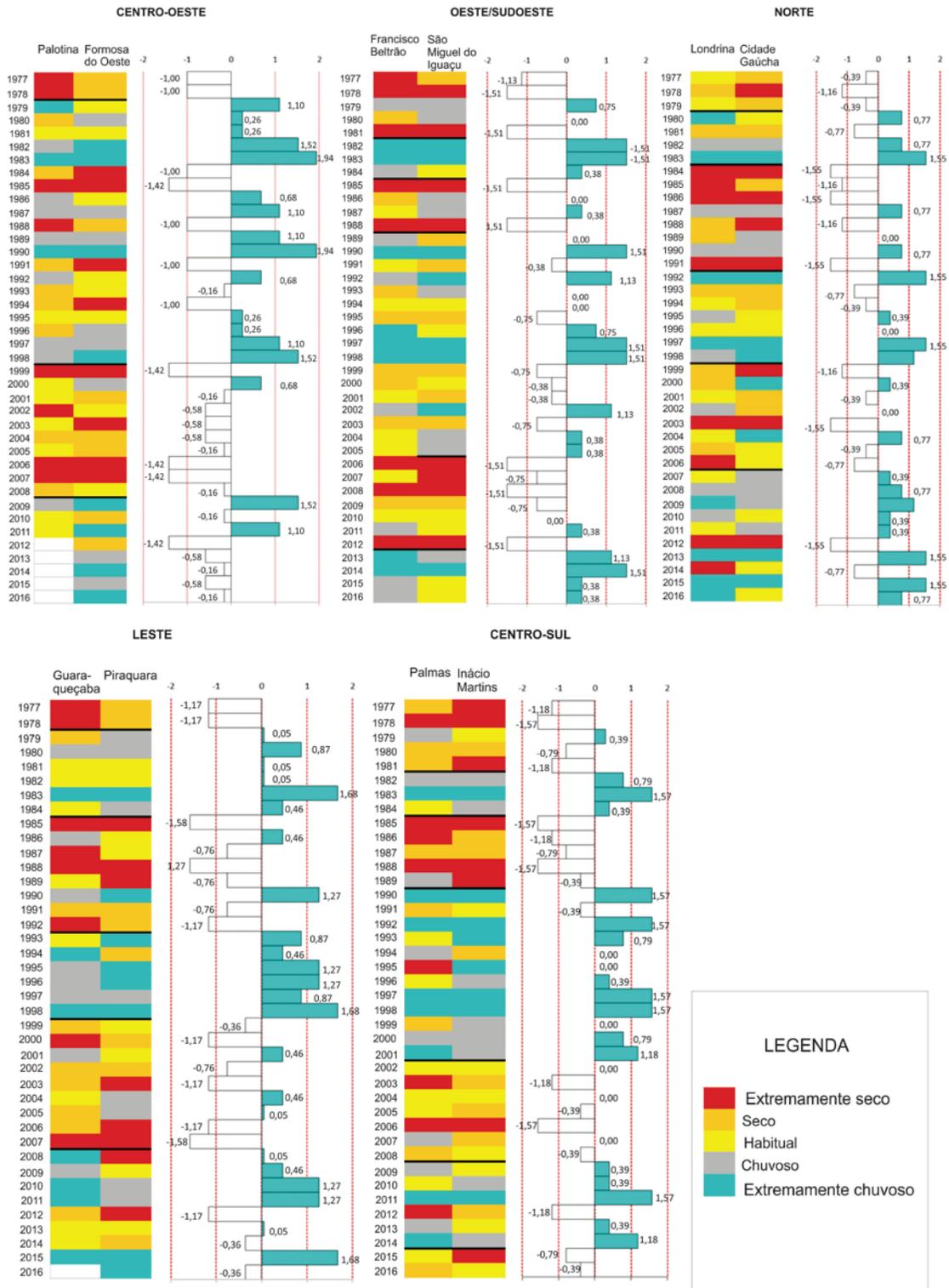


Figura 1 - Anos-padrões para cada região homogênea de precipitação para o Paraná.

O preenchimento de dados faltantes é realizado por meio de métodos estatísticos que utilizam medições realizadas em estações próximas, juntamente com equações matemáticas que relacionam estas medições com a estação que apresenta a ausência de dados (OLIVEIRA et al., 2010; BIER e FERRAZ, 2017). Geralmente, esses métodos consideram diferentes fatores com distintos pesos em seus cálculos, mas, frequentemente, a distância é o principal deles. O presente artigo enfoca as ausências de dados pluviométricos, que apresentam variações espaço temporais e que podem apresentar maiores ou menores semelhanças dependendo da proximidade ou do afastamento entre os postos de coleta. Portanto, considerando a relação da distância entre os postos pluviométricos disponíveis foram executados os seguintes métodos: Média aritmética, Ponderação Regional, Inverso da Distância e Inverse Distance Weithtning (IDW).

Ressalta-se que na aplicação dos referidos métodos foram utilizados cálculos da distância entre as estações, atribuindo ou não um peso para a influência deste fator.

A média aritmética é um método simples, que leva em consideração os dados das estações próximas para estimar um determinado valor ausente na estação em análise. Este método foi realizado considerando as cinco estações de maior proximidade da estação em análise. A média aritmética é a razão do somatório dos valores interpolados (Equação 1):

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

em que,  $D_x$  é o valor estimado;  $D_i$  o valor da estação vizinha e  $n$  é o número de estações próximas consideradas.

O método da ponderação regional estabelece regressões lineares entre os dados dos três postos mais próximos daquele que necessita do preenchimento de falhas, incorporando a informação da distância. Para cada uma das regressões lineares efetuadas obtém-se o coeficiente de correlação para que o dado seja estimado (OLIVEIRA et al, 2010; PRUSKI et al, 2004; MELLO, KOHLS e OLIVEIRA, 2017)

Na aplicação do método da ponderação regional, de acordo com Tucci (2001), foram considerados os dados diários de três postos com maior proximidade daquele que se desejava preencher os dados. Este método é usualmente aplicado para o preenchimento de falhas de precipitação em escalas mensais e anuais, mas optou-se por testar seu desempenho no preenchimento de dados diários de precipitações, conforme a Equação 2:

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

em que,  $D_x$  é o valor estimado;  $N_x$  e  $N_i$  são as precipitações médias das estações em análise e vizinha, respectivamente (mm);  $D_i$  é a precipitação mensal ou anual da estação vizinha no mesmo período em que a falha foi verificada e  $n$  é o número de estações vizinhas consideradas.

O inverso da distância (Equação 3) é um método que utiliza em sua estimacão a distância euclidiana entre a estação teste e as demais estações. Este método se assemelha ao Inverse Distance Weithtening (IDW), mas não atribui peso às distâncias das estações.

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

Segundo Silva et al. (2008), o método Inverse Distance Weithtening (IDW) é um método determinístico que utiliza as distâncias na integralidade das estações meteorológicas que compõem a base de dados, estabelecendo pesos para a influência que cada uma das estações tem em relação a estação que necessita de preenchimento dos dados. Este peso é determinado conforme a distância euclidiana entre as estações, tendo como critério a semelhança nos dados de estações que são mais próximas daquela que contém falhas, conforme a Equação 4:

$$D_x = \frac{\sum_{i=1}^n \left( \frac{1}{d_i^2} D_i \right)}{\sum_{i=1}^n \left( \frac{1}{d_i^2} \right)} \quad (4)$$

em que  $D_x$  é o valor a ser estimado;  $D_i$  o valor da  $i$ -ésima estação vizinha;  $d_i$  é a distância euclidiana e  $n$  é o número de amostragens.

### III. RESULTADOS E DISCUSSÃO

Os dados pluviométricos diários foram suprimidos dos anos selecionados das dez estações meteorológicas que compõem o presente estudo. Estas lacunas foram submetidas aos cálculos dos métodos

estatísticos descritos que, após o seu preenchimento, os dados estimados foram comparados com os dados observados nas respectivas estações meteorológicas.

Geralmente, a ausência de dados possui uma distribuição aleatória. Mas, não se pode confundir as lacunas com a não ocorrência de precipitação, que compreende dias em que não há registro de precipitações, ou seja, inexistência de chuva. Neste aspecto, a dispersão dos dados mostra que a mediana se concentra próximo a zero, pois a maioria dos dados compreendem dias em que não há registro de precipitações, ou seja, inexistência de chuva; bem como os quartis, em virtude de sua frequência (Figura 2).

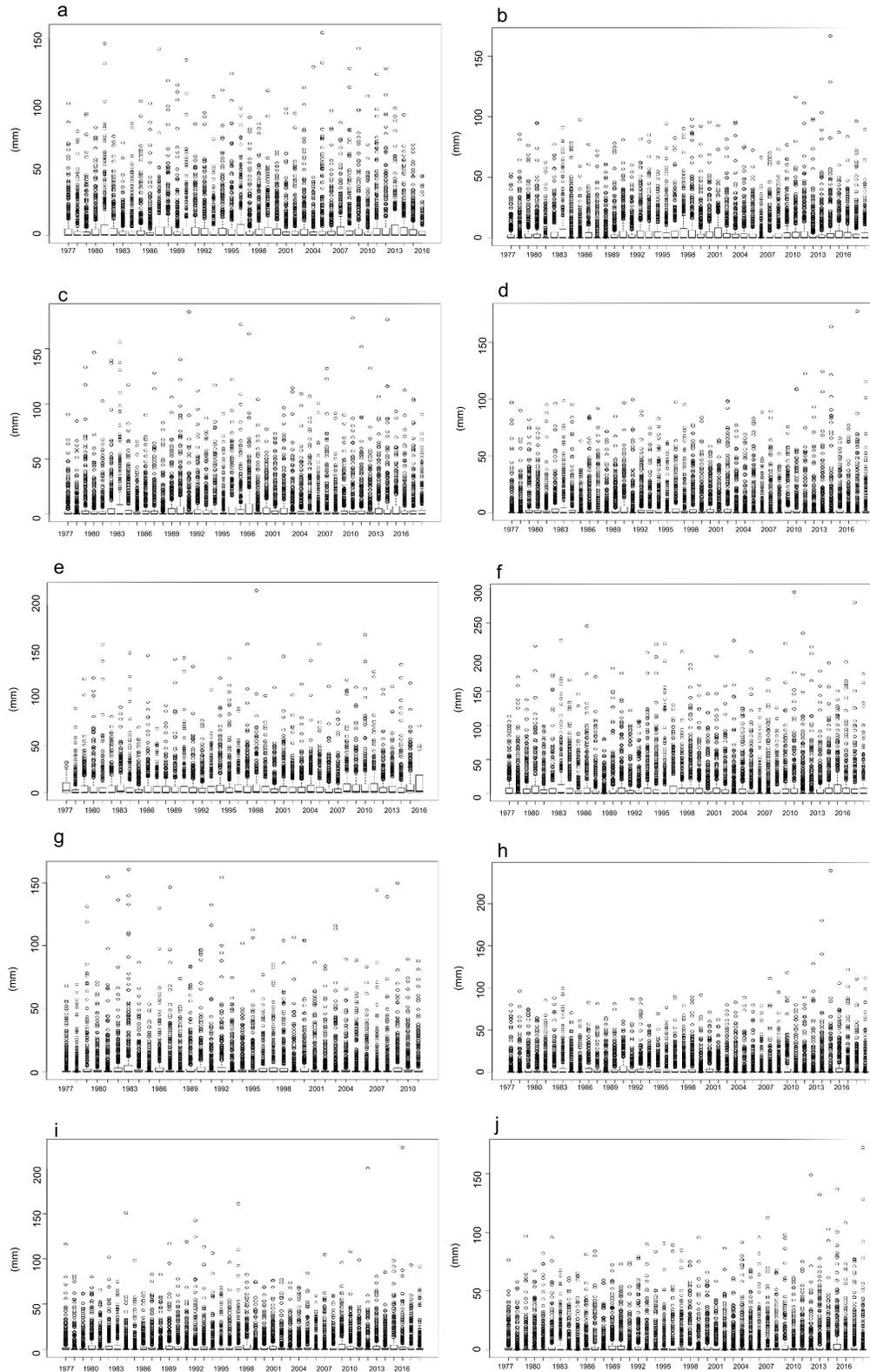


Figura 2 – Dispersão dos dados diários nas estações de referência. a) Palmas; b) Inácio Martins; c) Francisco Beltrão; d) São Miguel do Iguçu; e) Guaqueçaba; f) Piraquara; g) Palotina; h) Formosa do Oeste; i) Londrina; j) Cidade Gaúcha

Nas estações de Cidade Gaúcha, Londrina, São Miguel do Iguçu, Palotina e Formosa do Oeste frequentemente ocorrem precipitações com até 50mm diários. Em Inácio Martins, Piraquara e Francisco Beltrão são registradas precipitações mais frequentes, porém menos concentradas, enquanto em Guaraqueçaba e Palmas as chuvas são frequentes e com volumes maiores.

As Tabelas 1 a 5 apresentam a estatística descritiva para os dados considerados para as regiões Centro-Sul, Oeste/Sudoeste, Leste, Centro-Oeste e Norte, respectivamente. As classificações de cores representam a interpretação de melhores e piores ajustes dos métodos estatísticos, sendo que o azul indica a melhor aproximação do valor estimado ao observado e vermelho o valor mais distante do observado.

Em Palmas, no ano de 2002 (Tabela 1), o valor médio diário da precipitação observada foi de 5,54 mm e o método que produziu valores mais próximos foi a Média Aritmética (5,43 mm), enquanto os valores mais discrepantes foram obtidos com a Ponderação Regional. Em relação ao valor máximo observado (63,8 mm), o método que produziu um resultado mais próximo foi a Ponderação Regional (64,7 mm), assim como foi o método que registrou o desvio padrão com maior proximidade. No ano de 2004 a maior aproximação com a média observada (4,56 mm) foi a estimada pelo método IDW (4,55 mm), enquanto a Ponderação Regional estimou valores mais próximos para o máximo e o desvio padrão.

Na estação de Inácio Martins o valor médio observado no ano de 2002 foi 5,03 mm e os valores estimados pela Média Aritmética se ajustaram de forma mais próxima aos observados, bem como para a máxima e o desvio padrão (81,4 e 11,55 mm, respectivamente). Em 2004 os valores melhor estimados para a média e a máxima foram aqueles obtidos pelo Inverso da Distância e pela Média Artimética, respectivamente.

Tabela 1 – Estatística descritiva das precipitações (mm) na região Centro-Sul do estado do Paraná para os anos de 2002 e 2004. Obs.:

Anos	Métodos	Palmas				Inácio Martins			
		Méd.	Máx.	Min.	D. P.	Méd.	Máx.	Min.	D. P.
2002	Obs.	5,54	63,8	0,0	11,36	5,03	81,4	0,0	11,55
	IDW	5,05	49,7	0,0	8,15	4,86	57,4	0,0	8,63
	Inv. Dist.	4,74	48,6	0,0	7,63	4,70	51,0	0,0	7,85
	Média	5,43	55,0	0,0	8,70	4,95	71,1	0,0	10,03
	Pond. Reg.	6,33	64,7	0,0	11,57	4,64	67,0	0,0	9,62
2004	Obs.	4,56	93,0	0,0	11,58	4,65	75,5	0,0	10,75
	IDW	4,55	47,5	0,0	7,95	4,35	57,9	0,0	8,31
	Inv. Dist.	4,42	42,1	0,0	7,59	4,39	46,2	0,0	7,74
	Média	5,02	56,2	0,0	8,72	4,28	70,2	0,0	9,31
	Pond. Reg.	5,06	68,8	0,0	10,75	3,97	66,7	0,0	9,02

Observado; IDW: Inverso Quadrado da Distância; Inv. Dist.: Inverso da Distância; Pond. Reg.: Ponderação Regional; Méd.: Média; Máx.: Máxima; Min.: Mínima; D. P.: Desvio Padrão. Valores marcados com azul (vermelho) indicam a maior (menor) aproximação ao valor observado.

Observa-se na Tabela 2, para o ano de 1994, que as estações de Francisco Beltrão e São Miguel obtiveram valores estimados pelo método Inverso da Distância mais próximos aos valores médios observados. Contudo, o valor máximo estimado pelo método IDW em Francisco Beltrão obteve melhor aproximação com o observado, enquanto em São Miguel a Média Aritmética demonstrou melhor desempenho. Para estas estações, as melhores variações estimadas (D. P.) foram aquelas da Ponderação Regional e Média Aritmética. Para o ano de 2010 os melhores valores estimados em relação à média observada foram os obtidos por meio da Ponderação Regional, na estação de Francisco Beltrão e do Inverso da Distância na estação de São Miguel do Iguaçu. O método IDW estimou com melhor desempenho os valores máximos e desvio padrão na estação Francisco Beltrão, enquanto a Ponderação Regional foi mais eficaz no cálculo do valor máximo na estação de São Miguel do Iguaçu.

Tabela 2 – Estatística descritiva das precipitações (mm) na região Oeste/Sudoeste do estado do Paraná para os anos de 1994 e 2010.

Anos	Métodos	Francisco Beltrão				São Miguel do Iguaçu			
		Méd.	Máx.	Min.	D. P.	Méd.	Máx.	Min.	D. P.
1994	Obs.	6,27	117,0	0,0	15,04	4,8	82,9	0,0	12,44
	IDW	4,9	92,7	0,0	12,54	5,08	59,8	0,0	8,67
	Inv. Dist.	5,0	67,7	0,0	9,62	4,82	54,2	0,0	7,64
	Média	7,54	66,3	0,0	12,66	5,36	71,2	0,0	9,74
	Pond. Reg.	8,75	78,4	0,0	14,17	4,43	59,9	0,0	8,5
2010	Obs.	5,75	178,0	0,0	16,07	4,69	108,9	0,0	12,86
	IDW	4,81	113,7	0,0	13,82	4,95	57,9	0,0	8,69
	Inv. Dist.	4,79	69,2	0,0	10,43	4,83	53,1	0,0	7,85
	Média	5,06	78,7	0,0	12,45	5,13	59,3	0,0	9,45
	Pond. Reg.	5,12	80,3	0,0	12,58	4,43	59,6	0,0	8,44

Obs.: Observado; IDW: Inverso Quadrado da Distância; Inv. Dist.: Inverso da Distância; Pond. Reg.: Ponderação Regional; Méd.: Média; Máx.: Máxima; Min.: Mínima; D. P.: Desvio Padrão. Valores marcados com azul (vermelho) indicam a maior (menor) aproximação ao valor observado.

Na estação de Guaraqueçaba (Tabela 3) as melhores estimativas foram aquelas obtidas para a média, máxima e desvio padrão, como da Ponderação Regional para os anos de 1981 e 1982; com exceção da média no segundo ano analisado, melhor ajustada ao IDW. No ano de 2013 o melhor ajuste obtido foi o método da Média Aritmética. Entretanto, para a estação de Piraquara (Tabela 3) prevaleceu o método IDW para o ano de 1981, exceto para o desvio padrão que se ajustou melhor com o método Inverso da Distância. Para os anos de 1982 e 2013 o valor médio calculado com a Ponderação Regional obteve o melhor desempenho e valores máximos e desvio padrão com a Média Aritmética.

Tabela 3 – Estatística descritiva das precipitações (mm) na região Leste do estado do Paraná para os anos de 1981, 1982 e 2013.

Anos	Métodos	Guaraqueçaba				Piraquara			
		Méd.	Máx.	Min.	D. P.	Méd.	Máx.	Min.	D. P.
1981	Obs.	6,56	157,7	0,0	16,78	3,32	39,60	0,0	7,45
	IDW	4,07	46,85	0,0	7,06	3,43	40,20	0,0	5,88
	Inv. Dist.	4,06	52,54	0,0	6,96	3,83	50,40	0,0	6,50
	Média	4,58	65,87	0,0	8,97	3,14	34,80	0,0	5,87
	Pond. Reg.	5,43	70,15	0,0	11,24	2,72	41,30	0,0	5,74
1982	Obs.	6,61	81,0	0,0	12,9	3,83	69,80	0,0	9,60
	IDW	5,37	53,09	0,0	8,61	4,94	46,30	0,0	8,15
	Inv. Dist.	5,25	48,13	0,0	8,21	5,06	43,20	0,0	7,86
	Média	4,92	81,77	0,0	9,24	5,15	52,80	0,0	9,12
	Pond. Reg.	7,95	80,38	0,0	13,93	4,61	48,80	0,0	8,72
2013	Obs.	6,74	113,0	0,0	13,69	3,38	71,70	0,0	8,97
	IDW	5,56	73,47	0,0	9,13	4,78	55,90	0,0	8,24
	Inv. Dist.	5,4	58,03	0,0	8,77	5,09	51,80	0,0	8,36
	Média	5,78	124,2	0,0	11,09	4,26	56,60	0,0	8,10
	Pond. Reg.	2,59	52,81	0,0	5,97	4,17	51,90	0,0	8,43

Obs.: Observado; IDW: Inverso Quadrado da Distância; Inv. Dist.: Inverso da Distância; Pond. Reg.: Ponderação Regional; Méd.: Média; Máx.: Máxima; Min.: Mínima; D. P.: Desvio Padrão. Valores marcados com azul (vermelho) indicam a maior (menor) aproximação ao valor observado.

Com relação à estatística descritiva elaborada para as estações de Palotina e Formosa do Oeste (Tabela 4) identifica-se que, para os anos de 1981 e 1995, em Palotina, as melhores estimativas dos valores faltantes foram obtidas com o método do Inverso da Distância. Porém, os valores máximos foram melhor estimados pela Média Aritmética, bem como o desvio padrão. Em uma comparação com os dados observados na estação Formosa do Oeste pode-se dizer que as melhores imputações dos valores das precipitações máximas e desvio padrão foram aquelas obtidas com o método IDW para os anos 1981 e 1995. Mas vale ressaltar que, em média, os melhores cálculos para os anos de 1981 e 1995 foram aqueles da Média Aritmética e Ponderação Regional, respectivamente.

Tabela 4 – Estatística descritiva das precipitações (mm) na região Centro-Oeste do estado do Paraná para os anos de 1981 e 1995.

Anos	Métodos	Palotina				Formosa do Oeste			
		Méd.	Máx.	Min.	D. P.	Méd.	Máx.	Min.	D. P.
1981	Obs.	4,41	154,6	0,0	14,08	4,72	85,5	0,0	11,81
	IDW	4,06	51,10	0,0	7,07	4,06	76,2	0,0	10,02
	Inv. Dist.	4,07	53,8	0,0	6,99	4,19	57,5	0,0	7,61
	Média	3,16	64,9	0,0	7,66	4,48	68,0	0,0	9,62
	Pond. Reg.	1,03	23,9	0,0	2,65	3,79	57,0	0,0	8,41
1995	Obs.	4,52	112,7	0,0	12,61	4,64	76,9	0,0	11,39
	IDW	4,88	50,6	0,0	8,22	6,51	66,7	0,0	12,0
	Inv. Dist.	4,66	44,5	0,0	7,67	4,98	40,3	0,0	8,16
	Média	4,97	106,4	0,0	12,66	4,86	62,94	0,0	4,93
	Pond. Reg.	1,68	36,0	0,0	4,28	4,57	54,68	0,0	8,8

Obs.: Observado; IDW: Inverso Quadrado da Distância; Inv. Dist.: Inverso da Distância; Pond. Reg.: Ponderação Regional; Méd.: Média; Máx.: Máxima; Min.: Mínima; D. P.: Desvio Padrão. Valores marcados com azul (vermelho) indicam a maior (menor) aproximação ao valor observado.

Para as estações de Londrina e Cidade Gaúcha (tabela 5) as estimativas dos valores máximos e desvio padrão para os anos analisados se ajustaram melhor com o cálculo da Ponderação Regional. Contudo, no ano de 1995, em média, o método do Inverso da distância e o IDW tiveram melhores ajustes. Para o ano de 1996 a média nas duas estações estimada pela Média Aritmética ficou mais próxima da observada.

Tabela 5 – Estatística descritiva das precipitações (mm) para a região Norte do estado do Paraná para os anos de 1995 e 1996. Obs.:

Anos	Métodos	Londrina				Cidade Gaúcha			
		Méd.	Máx.	Min.	D. P.	Méd.	Máx.	Min.	D. P.
1995	Obs.	4,72	68,8	0,0	11,94	3,73	90,6	0,0	10,08
	IDW	4,26	43,6	0,0	7,86	4,42	60,96	0,0	8,16
	Inv. Dist.	4,56	43,61	0,0	7,60	4,49	44,37	0,0	7,52
	Média	4,33	50,6	0,0	8,63	4,49	81,22	0,0	9,67
	Pond. Reg.	4,27	59,98	0,0	9,06	4,55	84,23	0,0	9,83
1996	Obs.	4,53	96,1	0,0	11,38	3,98	89,20	0,0	10,74
	IDW	4,96	51,77	0,0	7,93	4,35	39,05	0,0	6,89
	Inv. Dist.	5,15	38,32	0,0	7,22	4,73	30,86	0,0	6,60
	Média	4,86	40,06	0,0	8,30	4,13	47,0	0,0	7,69
	Pond. Reg.	5,03	69,69	0,0	9,73	4,24	50,57	0,0	8,28

Observado; IDW: Inverso Quadrado da Distância; Inv. Dist.: Inverso da Distância; Pond. Reg.: Ponderação Regional; Méd.: Média; Máx.: Máxima; Min.: Mínima; D. P.: Desvio Padrão. Valores marcados com azul (vermelho) indicam a maior (menor) aproximação ao valor observado.

A identificação do melhor método para o preenchimento de falhas para o conjunto de dados analisado foi realizada por meio dos seguintes cálculos estatísticos: correlação linear, raiz do erro médio quadrático (RMSE), coeficiente de determinação ( $R^2$ ), coeficiente de determinação ajustado ( $R^2$  ajustado), critério de Akaike (AIC) e Bayesiano (BIC). Na Tabela 6 são apresentados os valores da correlação linear, calculada para os diferentes métodos aplicados para estimar os dados faltantes. Dos anos suprimidos considerados separadamente para as dez estações de referência somam 22 casos, 12 deles (54,5%) apresentaram o maior valor de correlação linear entre os observados e estimados pelo método estatístico IDW, seguido da Média Aritmética (6 vezes ou 27%), do Inverso da Distância (3 vezes ou 13,6%) e da Ponderação Regional (1 vez ou 4,5%).

Tabela 6 – Correlação linear entre os dados observados e os dados estimados pelos diferentes métodos aplicados para as estações meteorológicas e anos selecionados.

	Estações	Ano	IDW	Inver. Dist.	Média	Pond. Reg.	
Região Centro-Sul	IAPAR – Palmas	2002	<b>0,7279</b>	0,6527	0,6918	0,6887	
		2004	<b>0,7902</b>	0,7217	0,7509	0,7806	
	ANA – Inácio Martins	2002	0,8221	0,7710	0,8208	<b>0,8408</b>	
		2004	<b>0,7406</b>	0,6931	0,7233	0,7391	
Região Oeste/Sudoeste	IAPAR – Francisco Beltrão	1994	0,4958	0,6218	<b>0,6726</b>	0,6217	
		2010	0,4769	0,6154	<b>0,6381</b>	0,6312	
	ANA – São Miguel do Iguçu	1994	<b>0,8014</b>	0,7372	0,7498	0,7024	
		2010	<b>0,8528</b>	0,7889	0,8266	0,755	
	Região Leste	IAPAR – Guaraqueçaba	1981	0,3119	0,2969	<b>0,4209</b>	0,2457
			1982	<b>0,44</b>	0,4225	-0,0245	0,4164
Região Centro-Oeste	IAPAR – Palotina	2013	0,3761	0,3223	<b>0,5386</b>	0,179	
		1981	<b>0,6364</b>	0,5872	0,5713	0,5148	
		1995	0,5185	<b>0,6029</b>	0,3268	0,327	
	ANA – Formosa do Oeste	1981	0,4686	0,749	<b>0,7743</b>	0,7378	
		1995	0,4617	<b>0,6663</b>	0,6199	0,6407	
		1995	<b>0,7793</b>	0,6761	0,7205	0,7224	
Região Norte	IAPAR – Londrina	1996	<b>0,5384</b>	0,5219	0,5333	0,4559	
		1995	<b>0,7509</b>	0,7327	0,707	0,7158	
	ANA – Cidade Gaúcha	1995	<b>0,7509</b>	0,7327	0,707	0,7158	
		1996	<b>0,6696</b>	0,6109	0,6155	0,5761	

Valores em negrito indicam o maior valor de correlação linear para o ano-padrão em estudo.

Os maiores valores de correlação linear apresentados na Tabela 6 foram identificados nos postos de São Miguel do Iguçu ( $r=0,8528$ ) em 2010 e Inácio Martins ( $r=0,8408$ ) em 2002. Já os menores valores foram registrados em Guaraqueçaba para os três anos analisados. Esta estação obteve resultados díspares das demais e essa dificuldade dos métodos estatísticos em estimarem os dados faltantes pode estar relacionada à localização da estação na região litorânea, já que este fator contribui para chuvas localizadas, não semelhantes às estações vizinhas.

Na Figura 3 são apresentados os valores observados e os resultantes da aplicação dos métodos de preenchimento de falhas adotados para a estação de São Miguel do Iguçu no ano de 2010, que apresentou os maiores valores de correlação linear entre o observado e os métodos de imputação calculados (Tabela 6). Constata-se que os dados estimados acompanham o movimento da série, mas com uma menor amplitude,

---

indicando uma subestimação dos valores diários, replicando com maior precisão as menores precipitações diárias ou os dias em que não ocorre precipitação. Para eventos isolados, com precipitações diárias superiores a 20 mm, os métodos subestimam os valores, assim como para os eventos com precipitações diárias superiores a 60 mm. Mas quando essas precipitações ocorrem em dias consecutivos, os métodos produzem resultados similares aos observados (real). Com maior frequência a Ponderação Regional, Média Aritmética e Inverso da Distância superestimam as precipitações diárias.

---

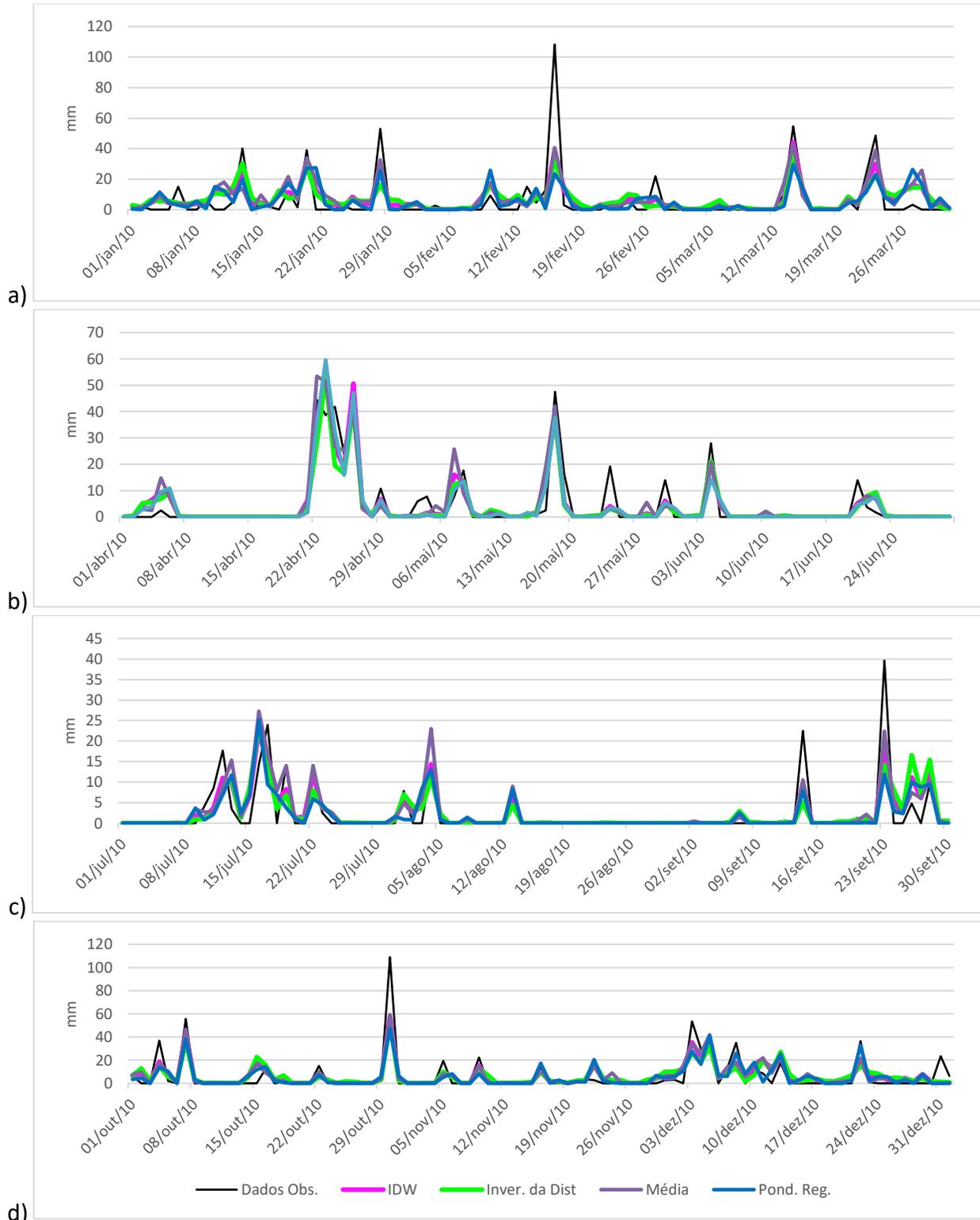


Figura 3 - Comparação dos valores observados e estimados no ano de 2010, na estação de São Miguel do Iguçu, sendo a) de janeiro a março; b) de abril a junho; c) de julho a setembro; d) de outubro a dezembro. Os gráficos não respeitam a mesma escala no eixo vertical para favorecer a visualização.

---

Para o posto de São Miguel do Iguaçu (2010) (Figura 3) o método IDW produziu dados que, mais frequentemente, se aproximam dos valores observados acima de 20mm. A média aritmética e a ponderação regional, no entanto, superestimam os valores das precipitações diárias maiores que 20mm.

Na Figura 4 estão representadas as precipitações diárias dos anos de 1981 em Guaraqueçaba, que possui a menor correlação linear entre os dados observados e estimados. Na Figura 5 percebe-se que os métodos não conseguiram estimar os eventos com precipitações diárias elevadas que ocorreram, principalmente, nos primeiros meses do ano (Figura 5a), enquanto nos meses mais próximos ao final do ano (Figuras 5c e 5d) os métodos produziram valores maiores do que os observados.

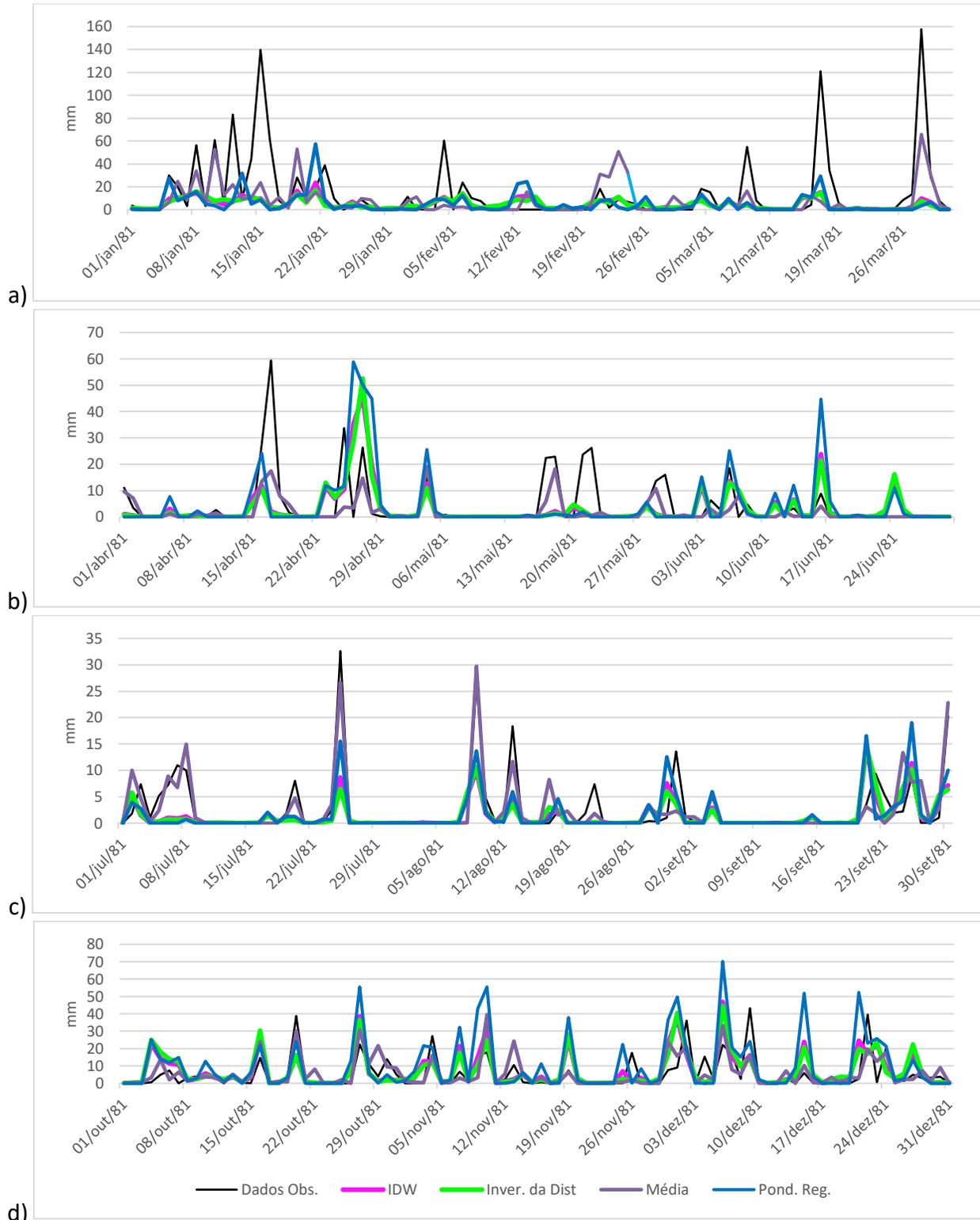


Figura 4 - Comparação dos valores observados e estimados no ano de 1981, na estação de Guaraqueçaba sendo a) de janeiro a março; b) de abril a junho; c) de julho a setembro; d) de outubro a dezembro. Os gráficos não respeitam a mesma escala no eixo vertical para favorecer a visualização.

Para o posto de Guaraqueçaba (Figura 4), localizado próximo ao litoral do Estado, onde as precipitações são mais frequentes, os métodos subestimam as precipitações diárias superiores a 40 mm.

#### IV. CONSIDERAÇÕES FINAIS

A aplicação dos diferentes métodos para o preenchimento das falhas nos dados pluviométricos diários para o estado do Paraná demonstrou que, para a variável média, a Média Aritmética produziu dados mais próximos dos observados; seguida do Inverso da Distância, Ponderação Regional e IDW. As precipitações máximas foram melhor estimadas pela Ponderação Regional, seguida da Média Aritmética e IDW. O desvio padrão obteve valores estimados mais próximos dos observados também pela Ponderação Regional, seguida da Média Aritmética e IDW.

Os métodos tiveram dificuldade para estimar os valores mínimos diferentes de 0mm. Também se conclui que os dados estimados acompanham a variabilidade da série, mas com uma menor amplitude, subestimando os valores diários, replicando com maior precisão as menores precipitações diárias ou os dias em que não ocorre precipitação. Para eventos isolados, com precipitações diárias superiores a 20 mm, os métodos subestimam os valores, assim como para os eventos com precipitações diárias superiores a 60 mm. Mas quando essas precipitações ocorrem em dias consecutivos, eles produzem resultados similares aos observados (real). Com maior frequência, os métodos Ponderação Regional, Média Aritmética e Inverso da Distância superestimam as precipitações diárias.

O método IDW é o que mais frequentemente produz resultados que se aproximam dos valores observados acima de 20mm. A Média Aritmética e a Ponderação Regional, no entanto, superestimam os valores das precipitações diárias maiores que 20mm.

Para os postos localizados próximos ao litoral do estado, onde as precipitações são mais frequentes, os métodos subestimam as precipitações diárias superiores a 40 mm.

Dos 22 casos (anos suprimidos considerados separadamente para as dez estações analisadas) 12 deles (54,5%) apresentaram o maior valor de correlação linear entre os dados observados e aqueles estimados pelo método estatístico IDW, seguido da Média Aritmética (6 vezes ou 27,%), do Inverso da Distância (3 vezes ou 13,6%) e da Ponderação Regional (1 vez ou 4,5%). Resultados que possibilitam definir o IDW como o método que apresentou melhor desempenho para estimar os dados das precipitações diárias faltantes em séries temporais para o estado do Paraná. Desempenho que pode ser justificado pelo fato deste método considerar

as distâncias euclidianas entre a integralidade das estações meteorológicas que compõem a base de dados e pelo estabelecimento de pesos que definem a influência que cada uma das estações têm em relação à estação que necessita de preenchimento dos dados; além do critério da semelhança entre os dados de estações que são mais próximas daquela que contém falhas

### Agradecimentos

Ao Núcleo de Inteligência Territorial do Parque Tecnológico de Itaipu (PTI) pelo financiamento da pesquisa e concessão dos dados.

### V. REFERÊNCIAS

- AIEB, A.; MADANI, K.; SCARPA, M.; LEFSIH, K. A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. *Heliyon*, 5(2019) e01247. doi: 10.1016/j.heliyon.2019.e01247
- 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.
- DUBREUIL, V.; MEROT, P.; DELAHAYE, D.; DESNOS, P. *Changement climatique dans l’Ouest*. Rennes: Presses Universitaires de Rennes, 2012.
- ELY, D. F. Padrões espaciais das tendências das precipitações sazonais e mensais no estado do Paraná (Brasil). *Revista Brasileira de Climatologia*, Ano 15, Edição Especial – XIII Simpósio de Climatologia Geográfica, jun/2019, p. 83 – 105.
- GARCÍA-PEÑA, M.; ARCINIEGAS-ALARCÓN, S.; BARBIN, D. Imputação de dados climáticos utilizando a decomposição por valores singulares: uma comparação empírica. *Revista Brasileira de Meteorologia*, v. 4, p. 527–536, 2014.
- LIMBERGER, L.; ELY, D. F. Régionalisation des précipitations dans l’état du Paraná (Brésil) em lien avec la température de surface de la mer. In: XXXII Colloque Internationale de l’Association Internationale de Climatologie. 2019, Thessaloniki. Actes [...], Thessaloniki, 2019. p. 493-498
- MELLO, Y. R. de; KOHLS, W.; OLIVEIRA, T. M. N. de Uso de diferentes métodos para o preenchimento de falhas em estações pluviométricas. *Boletim de Geografia*. Maringá, v. 35, n. 1, p. 112-121, 2017. DOI: <http://dx.doi.org/10.4025/bolgeogr.v35i1.30893> 113
- NOUACEUR, Z.; LAIGNEL, B.; TURKI, I. *Changement climatique en Afrique du nord: vers des conditions plus chaudes et plus humides dans le moyen Atlas marocain et ses marges*. In: Colloque de l’Association Internationale de Climatologie, n.27, 2014, Dijon. Actes [...]. Dijon, 2014. p. 387 – 393
- OLIVEIRA, L. F. C. de; FIOREZE, A.P.; MEDEIROS, A.M.M.; SILVA, M.A.S. Comparação de metodologias de preenchimento de falhas de série históricas de precipitação pluvial anual. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 14, n. 11, p. 1186–1192, 2010.
- PRUSKI, F.F.; PEREIRA, S.B.; NOVAES, L.F. de; SILVA, D.D. da; RAMOS, M.M. Precipitação média anual e vazão específica média de longa duração, na bacia do São Francisco. *Revista Brasileira de Engenharia Agrícola e*

Ambiental, Campina Grande, v. 8, n.2-3, 2004, p. 247-253.

TUCCI, C. E. M. Hidrologia: ciência e aplicação. 2. ed. Porto Alegre: ABRH, 2001.

VIEIRA, S. A.; OSORIO, D. M. M.; QUEVEDO, D. M. de; 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, ano 14, v.23, jul/dez, p. 189-204, 2018.

---