



## XXII SÍMPOSIO BRASILEIRO DE RECURSOS HÍDRICOS

26 de novembro a 01 de dezembro de 2017

Florianópolis- SC

### **Metodologias de preenchimento de falhas de séries pluviométricas anuais: Comparativo entre o método do Vetor Regional e a Regressão Linear Simples**

*Átila Alves Pinto*<sup>1</sup>; *Italo Ruan Dantas Ferreira*<sup>2</sup>; *Ticiano Marinho Carvalho Studart*<sup>3</sup> e *Renata Mendes Luna*<sup>3</sup> \*

**Resumo** – As séries históricas pluviométricas são essenciais na Hidrologia; entretanto, apresentam alguns períodos sem dados. Existem diversos métodos para preenchimento dessas falhas, do mais simples (Ponderação Regional) ao mais sofisticado (Vetor Regional). Estudos anteriores, realizados no Semiárido, apontaram o Método da Regressão Linear Simples como aquele que apresentou os melhores resultados. O objetivo deste trabalho é verificar se a maior complexidade do Vetor Regional incorre em resultados substancialmente melhores frente aos da Regressão Linear Simples. Utilizou-se 16 postos de quatro municípios do Estado do Ceará (Canindé, Quixadá, Quixeramobim e Sobral), inseridos em uma mesma região climática. Destes, selecionou-se um posto o qual teve 20%, 30%, 40% e 50% de seus dados propositalmente excluídos para simular falhas. No cálculo do Vetor Regional, utilizou-se o software HYDRACCESS, cujos resultados foram comparados com os obtidos pela Regressão Simples. Apesar de apresentar resultados levemente melhores, o Vetor Regional tem um maior grau de complexidade e demanda esforços mais elevados, sendo a diferença média dos desvios entre os dois métodos de apenas 4%. Portanto, para a região estudada, a Regressão Linear Simples é mais vantajosa pela simplicidade empregada.

**Palavras-Chave** – vetor regional, preenchimento de falhas, HYDRACCESS

### **Gap filling methodologies of annual rainfall series: Comparison between the Regional Vector method and Simple Linear Regression**

**Abstract** – Historical rainfall series are required for Hydrology; however this information often contains periods without data. There are several methods of gap filling, from the less complex (Regional Means) to the more sophisticated (Regional Vector). Previous studies, carried out in the

---

<sup>1</sup> Mestrando no Programa de Pós-Graduação em Eng. Civil – Recursos Hídricos (UFC): [atilaalves2010@hotmail.com](mailto:atilaalves2010@hotmail.com)

<sup>2</sup> Mestrando no Programa de Pós-Graduação em Eng. Civil – Recursos Hídricos (UFC): [italo.ruan70@gmail.com](mailto:italo.ruan70@gmail.com)

<sup>3</sup> Professoras do Programa de Pós-Graduação em Eng. Civil – Recursos Hídricos (UFC): Departamento de Engenharia Hidráulica e Ambiental - Centro de Tecnologia-Bloco 713 – Pici 60451-970 – Fortaleza/CE. E-mail: [ticianastudart2010@gmail.com](mailto:ticianastudart2010@gmail.com); [renata@deha.ufc.br](mailto:renata@deha.ufc.br)

Semi-arid region of Brazil, pointed to the Simple Linear Regression Method as the one that presented the best results. This study aims to verify if the greater complexity of the Regional Vector incurs better results than those of Simple Linear Regression. It was used 16 stations of four cities of the state of Ceará (Canindé, Quixadá, Quixeramobim and Sobral), located in the same climatic region. Of these stations, one was selected which had 20%, 30%, 40% and 50% of its data purposely excluded to simulate gaps. In the calculation of the Regional Vector, the software HYDRACCESS was used, whose results were compared with those obtained by Simple Regression. Despite of slightly better results, the Regional Vetor has a higher degree of complexity and demands higher efforts, with the average difference of deviations between the two methods being only 4%. Therefore, for the region studied, Simple Linear Regression is more advantageous for the simplicity employed.

**Keywords** – regional vector, gap filling, HYDRACCESS

## **INTRODUÇÃO**

O Sistema de Informações sobre Recursos Hídricos é um dos importantes instrumentos da Política Nacional dos Recursos Hídricos, este instrumento tem como função motivadora a coleta, tratamento e recuperação de informações sobre recursos hídricos visando à descentralização e produção de dados e informações, assim como garantir a toda sociedade o acesso aos mesmos. A Agência Nacional das Águas (ANA), autarquia federal com finalidade de implementar a Lei 9433/1997, disponibiliza as séries históricas através do Sistema de Informações Hidrológicas (“HidroWeb”), no qual podem ser coletadas várias informações sobre os postos pluviométricos e fluviométricos distribuídos pelo o país.

Os postos pluviométricos têm como objetivo obter, ao longo dos anos, uma série contínua de dados de precipitação. Todavia, por motivos de ausência das leituras pelo operador da estação ou até mesmo problemas no aparelho de registro, as estações pluviométricas podem apresentar falhas nas observações ou existência de períodos sem informações (ANA, 2011).

A presença de falhas nas séries históricas, possivelmente, conduzir-se-á a uma errônea interpretação dos resultados obtidos nos estudos considerados. A ausência de informações nas séries temporais, de fato, conduz a uma problemática que precisa ser levada em consideração pelos pesquisadores (Dantas *et. al*, 2016; Wanderley *et.al*, 2012).

Não é rara a frequência de falhas nesses postos, e, portanto, para que se possa aplicar tratamentos estatísticos às séries históricas, devem ser detectados e corrigidos os erros grosseiros para, em seguida, realizar o preenchimento das falhas existentes (Teegavarapu e Chandramouli, 2005).

Diversos métodos podem ser utilizados para o preenchimento de falhas de dados meteorológicos (Chibana *et. al.*, 2005), podendo incluir o uso de médias de dados observados ou dados obtidos de modelos geradores de dados.

Destacam alguns dos métodos, na literatura, empregados no preenchimento de dados - Método da Ponderação Regional, Método da Regressão Linear Simples e Método da Ponderação Regional com base em regressões lineares. Além destes, destaca-se o Método do Vetor Regional apresentando avanços em relação aos outros métodos, já que aquele realiza tanto o preenchimento de falhas quanto a análise de consistência dos dados pluviométricos mensais ou anuais. Ademais, por trabalhar com conceitos de máxima verossimilhança, ou seja, uma informação mais provável.

O Método Do Vetor Regional pode ser implementado sem maiores complicações em rotinas computacionais; o mesmo foi empregado para preenchimento de falhas, por exemplo, nos planos estaduais de recursos hídricos do Ceará e Piauí.

Oliveira *et al.* (2010) compararam os métodos mais usuais para preenchimento de falhas - ponderação regional, regressões linear e múltipla, ponderação regional com base em regressões lineares e vetor regional. Os autores, para um período de 22 anos de série histórica de seis estações pluviométricas localizadas no estado de Goiás, avaliaram os desvios relativos entre os valores reais e os estimados pelos métodos analisados. Concluiu-se que os menores desvios relativos foram apresentados pelo método da regressão linear múltipla; o pior desempenho ficou com o vetor regional. Entretanto, ocorreu uma melhora nos resultados deste quando se utilizou a combinação com os demais métodos empregados.

Leitão *et. al* (2016) realizaram uma avaliação comparativa entre os métodos da ponderação regional, regressão linear simples e ponderada para preenchimento de falhas em séries históricas de precipitação, para quatro municípios localizados na região semiárida do estado do Ceará, verificando que o método da regressão linear simples foi o que apresentou os menores erros relativo e médio.

### **Método da ponderação regional**

O método é visto como o mais simples e de maior utilização no preenchimento de dados faltosos em séries mensais e anuais, para tanto as precipitações do posto a ser completado são comparadas às de pelo menos mais três outros postos que estejam em uma mesma região climatológica e que possuam no mínimo dez anos de dados. A equação que rege o método é:

$$P_x = \frac{1}{3} \cdot \left( \frac{P_1}{P_{m1}} + \frac{P_2}{P_{m2}} + \frac{P_3}{P_{m3}} \right) \cdot P_{mx} \quad (1)$$

Onde:  $P_x$  = precipitação do posto X a ser determinada;  $P_1$ ,  $P_2$  e  $P_3$  = precipitações dos postos vizinhos;  $P_{m1}$ ,  $P_{m2}$  e  $P_{m3}$  = precipitações médias das estações vizinhas e  $P_{mx}$  = precipitação média do posto estudado.

### **Método da regressão linear simples**

Essa metodologia utiliza uma correlação entre o posto estudado e um outro vizinho que retorne o melhor resultado em comparação para o preenchimento das falhas utilizando a correlação linear múltipla:

$$P_x = aP_y + b \quad (2)$$

Onde:  $P_x$  = precipitação faltante do posto “x”;  $a$  = coeficiente angular da reta;  $P_y$  = precipitação do posto vizinho “y” e  $b$  = coeficiente linear da reta.

### Método da Ponderação Regional com base em regressões lineares

Diferentemente do método anterior, não somente um, mas mais de um posto são utilizados no preenchimento dos dados falhos, considerando uma ponderação entre eles através dos valores de  $R^2$ . A partir dos valores de  $R^2$ , calcula-se o peso de cada posto:

$$W_i = \frac{R_i^2}{\sum_{i=1}^n R_j^2} \quad (3)$$

Onde:  $W_i$  = peso do posto “i”;  $R_i^2$  = coeficiente de correlação do posto “i” e  $n$  = número de postos.

Assim, a precipitação é estimada por:

$$P_x = \sum_{i=1}^n W_i P_i \quad (4)$$

Onde:  $P_x$  = precipitação faltante no posto “x”;  $W_i$  = peso do posto “i”;  $n$  = número de postos e  $P_i$  = precipitação do posto “i”.

### Método do vetor regional

Desenvolvido por Hiez (1977) e Brunet-Monet (1979) é um método que, segundo Bertoni e Tucci (2007), consiste em uma séria cronológica, sintética de índices pluviométricos que tem origem em uma extração por meio do método estatístico de máxima verossimilhança das informações dos dados dos postos estudados. Sendo esses valores únicos para a região representados por dois vetores ótimos  $L$  (vetor regional coluna com “n” linhas) e  $C$  (vetor linha com “m” colunas) que formam uma matriz  $P$  de observações de  $m$  postos em  $n$  anos (ou meses). A estimativa de precipitação  $p$  no ano/mês “i” na estação “j” é dada por:

$$p_{ij} = L_i \cdot C_j \quad (5)$$

Para cada período existe um desvio que é utilizado no método dos mínimos quadrados assim minimizando a soma dos quadrados dos desvios. O processo torna-se iterativo e como primeiro chute,  $L_0$ , usa-se comumente a média das precipitações de todas as estações em cada período formando o vetor  $L$ . Com o chute inicial utiliza-se as seguintes equações para o cálculo dos vetores  $L$  e  $C$ :

$$C_j = \frac{\sum_{i=1}^n L_0 P_{ij}}{\sum_{i=1}^n (L_0)^2} \text{ com } j \text{ variando de } 1 \text{ a } m \quad (6)$$

$$L_i = \frac{\sum_{j=1}^m C_j P_{ij}}{\sum_{j=1}^m (C_j)^2} \text{ com } i \text{ variando de } 1 \text{ a } n \quad (7)$$

Bertoni e Tucci (2007) afirmam que a complexidade dos métodos acima apresentados, por ordem crescente são o Método da Ponderação Regional, Método da Regressão Linear Simples, Método da Ponderação Regional com base em regressões lineares e Método do Vetor Regional. O objetivo deste trabalho é, portanto, verificar o desempenho do método do Vetor Regional (o mais complexo) com o da Regressão Linear Simples, analisando se a complexidade daquele justifica o seu uso em relação aos resultados encontrados deste, para postos localizados na região semiárida do Ceará. Para tanto será utilizado o software “Hydraccess”, desenvolvido pelo Institut de Recherche pour le Developpement (IRD).

## **METODOLOGIA E DADOS**

Utilizou-se os postos estudados por Leitão et al (2016), localizados no interior do estado do Ceará, classificadas com o mesmo tipo climático. As estações escolhidas, quatro para cada caso, estão localizadas nos municípios de Sobral, Quixadá, Quixeramobim e Canindé.

Os dados foram obtidos no portal da ANA, “HidroWeb”, em formato txt e posteriormente importados para uma planilha Excel. Com a finalidade de estudar o erro provocado por cada método no preenchimento das falhas, uma porcentagem de dados foi excluída dos postos (20%, 30%, 40% e 50%).

No presente trabalho utilizamos uma base de dados chamada HYDRACCESS, desenvolvida por Philippe Vauchel, para o processamento do vetor regional. Dentro do programa, duas metodologias são apresentadas: a de Brunet-Moret (1979), baseada na média dos valores e a de Hiez (1977), que utiliza a moda das precipitações. Tra Bi (2013) afirma que os valores médios representam uma melhor estimativa por eliminar aqueles que são extremos; logo, a primeira metodologia foi a escolhida no presente trabalho.

Para a utilização do HYDRACCESS as informações de precipitação são inseridas em uma planilha com uma organização que permite a legibilidade dos dados. Após a leitura do arquivo, o mesmo retorna uma planilha com os valores calculados para cada estação estudada. Esses valores foram comparados com os valores reais e os desvios assim obtidos de dois diferentes modos: um erro médio para os desvios de cada ano e um outro na média geral dos anos.

Os dados após processamento no Hydraccess, foram analisados e comparados com os valores originais e um erro relativo foi calculado, o mesmo sendo dado por:

$$e_{rel} = \frac{|P_{real} - P_{est}|}{P_{real}} \quad (8)$$

$$\bar{e} = \frac{1}{m} \sum_{i=1}^m e_{rel,i} \quad (9)$$

Onde:  $e_{rel}$  = erro relativo para estimativa de cada ano;  $e_{rel,i}$  = erro relativo do ano  $i$ ;  $\bar{e}$  = erro médio;  $m$  = número de postos.

## RESULTADOS

Os erros entre os valores que foram apagados e os calculados por cada método podem ser vistos na Tabela 1 a seguir, assim como o erro médio relativo nas médias anuais.

Tabela 1 – Erros médios relativos da média anual individual e das médias anuais

Posto	Erro médio relativo da média anual individual		Erro médio relativo nas médias anuais	
	VR	Regressão Simples	VR	Regressão Simples
Sobral 20%	14%	24%	1%	5%
Sobral 30%	13%	16%	2%	6%
Sobral 40%	15%	19%	7%	7%
Sobral 50%	14%	19%	4%	11%
Canindé 20%	12%	16%	1%	2%
Canindé 30%	12%	13%	2%	3%
Canindé 40%	12%	12%	3%	8%
Canindé 50%	14%	12%	6%	5%
Quixeramobim 20%	12%	14%	3%	3%
Quixeramobim 30%	12%	8%	3%	3%
Quixeramobim 40%	14%	9%	14%	8%
Quixeramobim 50%	13%	18%	6%	0,4%
Quixadá 20%	11%	13%	2%	1%
Quixadá 30%	10%	8%	2%	0,4%
Quixadá 40%	12%	13%	4%	4%
Quixadá 50%	14%	21%	5%	5%

Quando se observa os valores de erro médio relativo da média anual para o vetor regional vemos que eles são mais constantes do que aqueles da regressão linear, que tendem a variar bastante quando se muda a quantidade de dados faltosos. O vetor regional apresenta, de um modo geral, erros

menores do que a regressão simples, sendo melhor em 11 dos 16 casos, com algumas exceções, apesar do menor erro ter sido encontrado pelo último método, 8%, assim como o maior erro pontual, 21%.

Diferentemente do imaginado o aumento da porcentagem de falhas não aumentou o erro, como se pode ser visto na Tabela 1, por exemplo Sobral 20% teve um erro, no vetor regional, de 14%, enquanto Sobral 40%, com o mesmo método, teve 40%. Pode-se perceber também que o menor valor ocorreu nas situações de 30% de falha para todos os grupos de postos nos dois métodos.

Quando se usa como parâmetro o erro provocado na média dos valores anuais das precipitações, pode-se ver que esses são bem menores do que aqueles. Nesse caso o vetor regional se mostra melhor em 6 casos, obtém o mesmo resultado que a regressão simples em outros 5 e foi pior em 5. Assim, percebe-se a que a média dos valores anuais é menos afetada pelas falhas do que a média dos valores anuais individuais.

## **CONCLUSÕES**

Os erros calculados pelos dois métodos apresentam valores um pouco elevados, na ordem de mais de 10%, porém vê-se um melhoramento nesses resultados com a utilização do Vetor Regional. Além de apresentar uma maior uniformidade nos erros, sendo menos dependente do número de falhas. Ao analisar o erro relativo nas médias anuais, observa-se menores erros e também resultados mais equilibrados entre os dois métodos, com uma leve vantagem para o Vetor Regional. Tendo em vista alcançar melhores resultados, a utilização de séries mais longas pode favorecer essa melhoria, sendo essa uma sugestão para futuros trabalhos.

Apesar de apresentar resultados levemente melhores, o Vetor Regional tem um maior grau de complexidade e demanda esforços mais elevados, sendo a diferença média dos desvios entre os dois métodos de apenas 4%. Portanto, para a região estudada, a Regressão Linear Simples é mais vantajosa pela simplicidade empregada.

## **REFERÊNCIAS**

BERTONI, J. C.; TUCCI, C. E. M. Precipitação. In: TUCCI, C. E. M. Hidrologia: Ciência e Aplicação. UFRGS, Porto Alegre, p.177-241, 2007.

BRASIL. Lei nº 9.433 de 8 de janeiro de 1997. Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 da Constituição Federal, e altera o art. 1º da Lei nº 8.001, de 13 de março de 1990, que modificou a Lei nº 7.990, de 28 de dezembro de 1989. Diário Oficial da União, Brasília, 9 de jan. de 1997.

BRUNET-MORET, Y. Homogénéisation des précipitations. Cah. Orstom, série Hydrologie, col XVI, nº 3 et 4, 147-170, 1979.

CHIBANA, E. Y.; FUMIGNAN, D.; MOTA, R. G.; VIEIRA, A. S.; FARIA, R. T. Estimativa de falhas em dados meteorológicos. In: Congresso Brasileiro de Agroinformática, 9, 2005. Londrina. Anais...Londrina, ABI-AGRO, 2005.

DANTAS, Leydson Galvíncio; DOS SANTOS, Carlos Antonio Costa; DE OLINDA, Ricardo Alves. Reamostragem de Séries Pluviométricas no Estado da Paraíba. Revista Brasileira de Geografia Física, v. 9, n. 4, p. 997-1006, 2016.

HIEZ, G. Processamento dos dados pluviométricos do Nordeste: a homogeneização dos dados. Recife: SUDENE/ORSTOM, 1978.

HIEZ, Gérard. L'homogénéité des données pluviométriques. Cahiers ORSTOM, série Hydrologie, v. 14, n. 02, p. 129-173, 1977.

LEITÃO, F.T.P; GIORDANI, L.C; STUDART, T.M.C. Análise intercomparativa de métodos de preenchimento de falhas de precipitações anuais para o estado do Ceará. In Anais do XIII Simpósio de Recursos Hídricos do Nordeste, Aracaju, 2016.

OLIVEIRA, L. F. C. de; 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.

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. 191–206. 2005.

TRA BI, Z. A. Étude de l'impact des activités anthropiques et de la variabilité climathique sur la végétation et les usages des sols, par utilisation de la télédétection et des statistiques agricoles, sur le bassin versant du Bouregreg (Maroc). Thèse de Doctorat en cotutelle, Université Félix Houphouët Boigny d'Abidjan, 190 p, 2013.

WANDERLEY, H. S.; AMORIM, R. S. C.; CARVALHO, F. O. Variabilidade espacial e preenchimento de falhas de dados pluviométricos para o estado de Alagoas. Revista Brasileira de Meteorologia, v.27, p. 347-354, 2012.