



Geoestatística no estudo de modelagem temporal da precipitação

**André Q. de Almeida¹, Aristides Ribeiro¹, Yhasmin G. Paiva¹,
Nilton Jr. L. Rascon¹ & Evaldo P. Lima²**

RESUMO

A análise geoestatística é uma poderosa ferramenta utilizada em estudos de dependência espacial. No que tange à dependência temporal, poucas são as análises realizadas com essa metodologia. Neste trabalho foi utilizada a técnica de geoestatística para ajustar um modelo de série temporal de precipitação, cujo poder é avaliado em prever valores futuros. O estudo foi realizado na Bacia do Rio Itapemirim em uma série de precipitação mensal de 1940 a 2006. O modelo geoestatístico foi comparado com um modelo de séries temporais de Box e Jenkins. O modelo geoestatístico ajustado foi aquele com 96 vizinhos utilizados na previsão e erro relativo absoluto médio de 80,46. A metodologia de geoestatística apresentou melhores resultados na estimação da precipitação mensal, em relação ao modelo SARIMA $(2,1,1)(0,1,1)_{12}$.

Palavras-chave: Bacia do Rio Itapemirim, Box & Jenkins, series temporais

Geostatistics in the temporal modeling study of precipitation

ABSTRACT

Geostatistics analysis is a powerful tool which has been used in studies of spatial dependence. With respect to the temporal dependence, few analysis are performed with this kind of methodology. In this study, the geostatistics technique was used to adjust a model to a series of temporal values of precipitation and its performance to predict data was also evaluated. The study was carried out at the Rio Itapemirim Basin in a series of monthly precipitation from 1940 to 2006. The adjusted model was compared to the Box and Jenkins model. The best geostatistics model adjusted was the one which used with 96 neighbors in the prediction and mean relative absolute error of 80.46. Geostatistics methodology showed better results for estimating the monthly precipitation in relation to the SARIMA model $(2,1,1)(0,1,1)_{12}$.

Key words: Bacia do Rio Itapemirim, Box & Jenkins, time series

¹ DEA/UFV, Av. Peter Henry Rolfs s/n Campus Universitário, CEP 36570-000, Viçosa, MG. Fone: (31) 3899-1906. E-mail: andreqa@gmail.com; ribeiro@ufv.br; yhasminp@hotmail.com; niltonrascon@yahoo.com.br

² Embrapa CNPS, Rua Jardim Botânico 1024, Jardim Botânico, CEP 22460-000, Rio de Janeiro, RJ. Fone: (21) 2179-4536. E-mail: evaldo@cnps.embrapa.br

INTRODUÇÃO

Dentre os parâmetros meteorológicos, a precipitação é a que apresenta a maior variabilidade temporal e espacial sendo, portanto, o de maior dificuldade de estimativa. Com isto, tornam-se fundamentais a caracterização do seu padrão local e suas variações. A chuva pode ser considerada a principal forma de suprimento hídrico para as atividades humanas e econômicas, principalmente na agricultura, setor em que a precipitação é usada, direta ou indiretamente, em processos de tomada de decisão, como segue: no manejo da irrigação, no risco de incêndios, na previsão de geadas, no zoneamento agroclimático, na ocorrência de pragas e doenças, na classificação climática, entre outros.

Em diversos trabalhos já se analisou a dependência espacial da precipitação (Melo Júnior et al., 2006; Mello & Silva, 2009). Uma técnica comumente empregada nesse tipo de estudo, é a geoestatística, que parte do princípio de que há forte correlação entre observações no espaço. Além disso, possui ferramentas determinísticas e estatísticas que possibilitam entender e modelar a variabilidade espacial (Deutsch & Journel, 1998).

A geoestatística se tem mostrado de grande utilidade para caracterizar e mapear a variação espacial das propriedades dos sistemas (Couto et al., 1997). Ela consiste da variografia, krigagem, simulações e de estratégias que otimizem os planos de amostragem dos atributos (Webster & Oliver, 1990).

Acredita-se que, poderosas, também podem ser suas predições no domínio do tempo; são escassos ainda os estudos nos quais se utiliza a técnica de geoestatística na ótica temporal, comparativamente, por exemplo, ao que se passa no domínio espacial (Holawe & Dutter, 1999; Carvalho et al., 2009). Mingot & Neves (1999), apresentaram um exemplo de utilização da metodologia geoestatística aplicada à análise de séries temporais. Esses autores aplicaram a metodologia em outros estudos (Mingoti & Glória, 2002). Lima et al. (2008), estudaram a variabilidade temporal da precipitação mensal no município de Alegre, Sul do Estado do Espírito Santo.

Os modelos de Box e Jenkins (Box et al., 1994), são amplamente utilizados em estudos de séries temporais. Silva et al. (2008), estimaram a temperatura média mensal de Uberlândia, MG. Referida metodologia tem, como objetivo principal, ajustar modelos em séries de tempo e posteriormente realizar previsões. Estudos mostram que em modelos sazonais autorregressivos, médias móveis (SARIMA) se ajustam satisfatoriamente aos dados de precipitação (Ferraz et al., 1999).

O objetivo deste trabalho foi modelar uma série temporal de precipitação por meio de geoestatística e realizar previsões em pequenas escalas de tempo. Na próxima seção são apresentados os dados e as metodologias utilizadas, com uma breve descrição da teoria de Geoestatística e de Box e Jenkins.

MATERIAL E MÉTODOS

A Bacia Hidrográfica do Rio Itapemirim está localizada na região sul do Espírito Santo; com área total de aproximadamente

5.923 km², abrange 17 municípios. Segundo a classificação de Köppen, o clima se enquadra no tipo Cwa, com inverno seco e verão chuvoso.

As informações de precipitação são da estação pluviométrica de Rive (Cod. 02041003). Pertence à bacia do Atlântico, trecho Leste, sub-bacia do rio Itapemirim. A localização geográfica da estação é -20° 44' de latitude e -41° 27' de longitude. As informações estão disponíveis no site da Agência Nacional das Águas (ANA), na guia hidroweb (<http://hidroweb.ana.gov.br/>). Para o período de análise do ano de 1940 ao de 2006, a precipitação total anual média foi de aproximadamente 1350 mm.

A geoestatística é uma ferramenta de análise espacial baseada na técnica de Kriging (Krige, 1951), com a finalidade de estimar valores não amostrados. Utiliza o modelo probabilístico associado à localização espacial dos pontos amostrados possibilitando, desta forma, uma análise da variabilidade entre os valores observados. Está baseada na teoria das variáveis regionalizadas (Matheron, 1963) que parte da premissa de haver dependência espacial entre as observações vizinhas de uma variável aleatória qualquer, distribuída continuamente no espaço.

Neste estudo, no lugar do espaço será analisada a dependência na escala do tempo, o que será feito entre pares de observações subsequentes. Assim sendo, a distância h será substituída pelo intervalo de tempo t ; o modelo de semivariograma será dado pela Eq. 1, abaixo:

$$\gamma(t) = \frac{1}{2N(t)} \sum_{i=1}^{N(t)} [Z(x) - Z(x+t)]^2 \quad (1)$$

em que: $\gamma(t)$ é o semivariograma para um vetor mês t ; $Z(x)$ e $Z(x+t)$ são os pares de valores de precipitação separados por um intervalo de tempo (mês) e $N(t)$ é o número de pares de valores medidos.

A estimação dos dados é o passo seguinte da análise. Após ajuste e validação do semivariograma, é então executada a Krigagem ordinária e os valores encontrados são utilizados para predir valores não amostrados; nesse caso, valores futuros da série de tempo precipitação. O estimador utilizado é o da krigagem (Krige, 1951), como segue:

$$\hat{Z}(t) = \sum_{i=1}^{N(t)} \lambda_i Z(x_i) \quad (2)$$

em que: $t > x_n$ e $Z(t_1), Z(t_2), Z(t_n)$, representam, respectivamente, os valores observados da série ou os valores da vizinhança usada na previsão de $Z(t)$ e λ_t são os valores de Lambda estimados pela equação abaixo:

$$\lambda_{(0)} = \Gamma_0^{-1} \gamma_0 \quad (3)$$

onde:

$$\gamma_0 = (\gamma(-t_1), \gamma(s-t_2), \dots, \gamma(s-t_n), 1) \quad (4)$$

e é a matriz (n+1)x(n+1) dada por:

$$\Gamma_0 = \begin{cases} \gamma(t_i - t_j), i = 1, \dots, n, j = 1, \dots, n, \\ 1, i = n + 1, j = 1, \dots, n, \\ 0, i = n + 1, j = n + 1, \\ 1, i = 1, \dots, n, j = n + 1 \end{cases} \quad (5)$$

sendo que γ_0 representa o semivariograma do processo e λ é o vetor contendo os coeficientes $\lambda_i, i=1,2,\dots,n$.

De modo geral, toda série temporal ($Z_t, t = 1:n$) pode ser decomposta em uma componente de tendência, outra sazonal e uma aleatória, do tipo $Z_t = T_t + S_t + a_t$. A tendência (T_t) pode ser entendida como aumento ou diminuição gradual nos valores da série ao longo do período; a série com sazonalidade (S_t) apresenta repetição de comportamento ao longo do tempo e a_t , a componente aleatória ou erro. Espera-se que a_t seja um ruído branco com média zero e variância constante.

Segundo a metodologia de Box et al. (1994), dada uma série não-estacionária na média Z_t , comumente se utiliza o operador diferença ∇ para tornar a série estacionária. Para um inteiro $d \geq 0$, esse operador atua da forma apresentada na Eq. 6 abaixo:

$$\Delta^d Z_t = (1 - B)^d Z_t \quad (6)$$

onde B é o operador de atraso definido por $B^k Z_t = Z_{t-k}$.

Um processo Z_t diz-se um modelo autorregressivo integrado de médias móveis, abreviado por ARIMA(p, d, q) se, satisfaz,

$$\varphi(B)\Delta^d Z_t = \theta(B)a_t \quad (7)$$

em que

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

e

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

são polinômios em B de graus p e q , chamados, respectivamente, polinômios autorregressivos estacionários e médias móvel invertíveis.

Quando um padrão dinâmico de sazonalidade estocástica está presente na série, este pode ser modelado com pequena modificação sobre o modelo ARIMA, dado na Eq. 6. Esta modificação tem a forma dada:

$$\varphi(B)\Phi(B)^s \nabla^d \nabla_s^D Z_t = \theta(B)\Theta(B^s) a_t \quad (8)$$

em que d e D representam o número de diferenças simples e sazonal feitas sobre Z_t com o período sazonal s .

O modelo dado na Eq. 8 é denominado SARIMA (p,d,q)x(P,D,Q).

A consistência dos dados e a verificação de prováveis falhas nos mesmos, foram verificadas pelo programa Hidro, versão 1.2 da ANA.

A análise de geoestatística foi utilizada com o auxílio do programa computacional PrevGeo, que usa a metodologia de Geoestatística introduzida por Mingoti & Neves (1999).

Realizaram-se as previsões com diferentes números de vizinhos e se escolheu o modelo que apresentou melhores resultados estatísticos, com base no erro médio simples (EM) e no erro quadrático médio (EQM) e pelo erro absoluto relativo médio (ABSERREM) nas previsões de um passo à frente.

Os modelos de Box e Jenkins foram ajustados pela linguagem computacional estatística R, versão 2.0.1. O melhor modelo ajustado aos dados foi escolhido com base nas menores variâncias e nos menores valores de Critério de Informação de Akaike (AIC) (Akaike, 1974).

De posse do modelo de geoestatístico e de Box e Jenkins, as previsões de um passo à frente foram realizadas para o período de 24 meses; os resultados foram avaliados pelas estatísticas do erro médio simples (MSE), absoluto (MAE) e erro quadrático médio (RMS).

RESULTADOS E DISCUSSÃO

Na série original composta de 67 anos, verificou-se a existência de falhas em alguns meses. Os anos 1989 e 2001, foram retirados da análise. Para ajuste dos modelos foi considerada a série modificada dos anos 1940 a 2004 (Figura 1); os dois últimos (2005 e 2006) anos, foram utilizados para validação dos modelos.

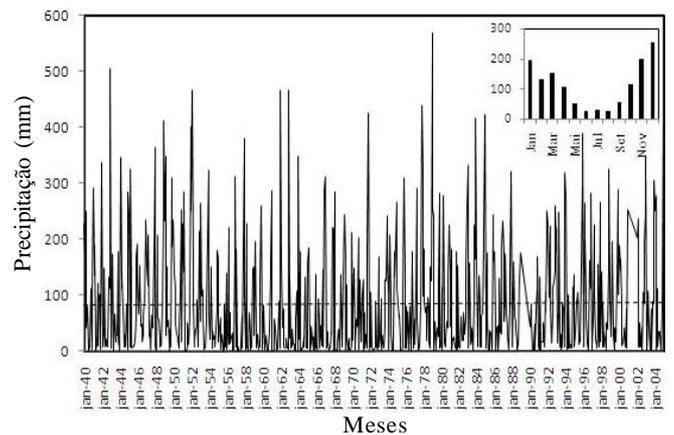


Figura 1. Série mensal modificada (sem os anos de 1989 e 2001) de precipitação para os anos de 1940 e 2004, seu valor médio e o padrão histórico

A precipitação média mensal da série da Figura 1 foi de 85,11 mm. Nota-se também a existência de outliers, com valores bem acima da média. Interessante notar, ainda, que o regime mensal das chuvas segue o padrão da classificação climática de Köppen, com inverno seco e verão chuvoso.

O melhor modelo geoestatístico ajustado para a série de precipitação foi aquele com 96 vizinhos (meses), ou seja, foram

utilizados 8 anos da série observada para realizar a previsão dos novos valores.

As estatísticas mostram que esse modelo obteve os menores erros, inclusive para o ABSERREM para um passo à frente de previsão. Os valores encontrados para o EM, EQM e o ABSERREM foram, respectivamente, 8,60; 73,90 e 80,46. A variância do erro de predição e seu desvio padrão, foram de 6060,91 e 77,85 respectivamente. A equação de predição de vizinhança 96 é como segue:

$$\hat{Z}(s) = \lambda_1 Z(s-1) + \lambda_2 Z(s-2) + \dots + \lambda_{95} Z(s-95) + \lambda_{96} Z(s-96) \quad (9)$$

Os respectivos valores de Lambda (λ) são mostrados na Tabela 1.

Tabela 1. Valores de Lambada (λ) ajustados para o modelo Geoestatístico com 96 vizinhanças (meses) de previsão*

0,028	-0,051	0,017	-0,037	-0,010
-0,004	-0,002	-0,005	0,009	-0,005
0,013	0,076	-0,015	-0,027	-0,062
0,039	0,042	-0,003	0,011	0,019
-0,012	-0,063	-0,069	0,045	0,036
0,033	0,061	0,036	0,027	0,058
0,007	-0,004	-0,014	-0,026	-0,007
0,009	0,055	0,068	0,099	0,097
0,033	0,013	-0,039	0,030	-0,028
0,050	0,047	0,079	0,123	0,087
0,025	0,013	0,021	0,007	0,022
0,008	0,038	0,069	0,043	0,078
0,095	0,047	-0,053	-0,002	-0,010
0,018	-0,015	-0,025	-0,017	-0,018
0,057	-0,019	0,029	0,040	-0,078
-0,025	-0,113	-0,065	-0,052	-0,017
0,020	0,031	0,026	0,032	0,064
0,076	-0,068	-0,044	0,014	-0,002
0,004	-0,006	-0,008	-0,052	-0,037
-0,016				

* Sequência dos valores (λ): da esquerda para a direita, indo de $\lambda = 1$ até $\lambda = 96$

Na modelagem de Box e Jenkins, as funções amostrais de autocorrelação e autocorrelação parcial indicam haver uma sazonalidade com período de 12 meses ($s=12$). Ante tal informação se sabe que o modelo final ajustado será um SARIMA de sazonalidade 12. Escolheu-se o modelo com menor valor de AIC (-8736,60), resultando em um SARIMA (2,1,1)(0,1,1)₁₂, após a identificação dos parâmetros do modelo, a equação de estimação de valores futuros ficou como segue:

$$\begin{aligned} (1 + 0,0051B + 0,0110B^2)(1 - B^{12})Z_t = \\ = (1 + 0,9817B)(1 + 0,9508B^{12})a_t \end{aligned} \quad (10)$$

Para verificar a adequação de um modelo, a teoria estatística de séries temporais faz propostas básicas das suas propriedades, tais como não correlação entre a normalidade dos erros (Wei, 1990). Nos modelos ajustados essas características foram avaliadas nos erros ajustados e, portanto, consideradas.

Os valores previstos por ambas as metodologias aparecem na Figura 2; em geral, o desempenho dos modelos foi satisfatório. Desconsiderando-se os meses com valores extremos como, por exemplo, dezembro, os valores preditos estão todos próximos do observado; nota-se também, que os valores estimados seguem a tendência dos dados originais, com máximas de precipitação no verão (período úmido) e mínimas no inverno (período seco). Sugere-se, para este mesmo estudo, a utilização de técnicas de modelagem que consideram, no ajuste dos modelos, os outliers podendo obter, assim, melhores resultados.

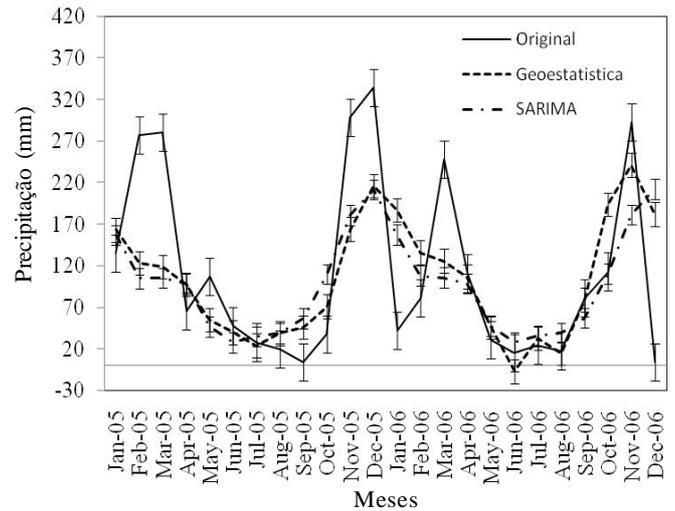


Figura 2. Valores observados (linha sólida), valores preditos por geoestatística (linha tracejada) e pelo modelo SARIMA (linha tracejada intercalada por ponto) para os 24 meses dos anos de 2005 e 2006

Apesar de apresentar melhores valores de estimação (Tabela 2), o modelo de geoestatística apresentou para o mês de junho de 2006, um valor de -7 mm de precipitação. Sabe-se que, neste caso, o valor predito deveria ser de 0 mm, ou próximo a isso. Ressalta-se, então, que uma análise de sensibilidade nos resultados deve ser feita, retirando-se os possíveis valores inconsistentes, que foram gerados.

Tabela 2. Estatística dos erros estimados pelos dois modelos ajustados, geoestatística e SARIMA

Estatística	Modelo	
	Geoestatística	SARIMA
Média	104,01	98,11
Desvio padrão	69,92	59,18
MSE	14.160,78	16.197,38
RMSE	119,00	127,27
MAE	122,67	131,33
r ²	0,40	0,32

Mesmo não sendo estatisticamente diferente dos erros apresentados pelo modelo SARIMA, a modelagem de Geoestatística apresentou os menores erros de previsão (Tabela 2). O teste de TUKEY, de médias não identificou diferenças nos resultados. Mingoti & Glória (2002), encontraram resultado semelhante ao comparar as duas metodologias.

CONCLUSÕES

1. O melhor modelo ajustado através da técnica de geoestatística, foi o com 96 vizinhos (meses) utilizados na predição da precipitação mensal;
2. Em geral, o modelo geoestatístico ajustado apresentou melhor desempenho na modelagem de previsão da precipitação mensal em relação ao modelo SARIMA.

LITERATURA CITADA

- Akaike, H. A new look at statistical model identification. *IEEE Transactions Automatic Control*, v.19, p.716-723, 1974.
- Box, G. E. P.; Jenkins, C. M.; Reinsel, G. C. *Time series analysis: Forecasting and control*, 3 ed., New York: Prentice-Hall, 1994.
- Carvalho, J. R. P.; Vieira, S. R.; Grego, C. R. Comparação de métodos para ajuste de modelos de semivariograma da precipitação pluvial anual. *Revista Brasileira de Engenharia Agrícola e Ambiental*. v.13, n.4, p.443-448, 2009.
- Couto, E. G.; Stein, A.; Klamt, E. Large area spatial variability of soil chemical properties in central Brazil. *Agriculture Ecosystems and Environment*, v.66, p.139-152, 1997.
- Deutsch, C. V.; Journel, A. G. *GSLIB – Geostatistical software library and user's guide*. 2.ed. Oxford: Oxford University Press, 1998. 369p.
- Ferraz, M. I. F.; Sáfiadi, T.; Lage, G. Uso de modelos de séries temporais na previsão de séries de precipitação pluviométricas mensais no município de Lavras, MG. *Revista Brasileira de Agrometeorologia*, v.7, p.259-267, 1999.
- Holawe, F.; Dutter, R. Geoestatistical study of precipitation series in Austria: time and pace. *Journal of Hydrology*, v.219, p.70-82, 1999.
- Krige, D. G. A statistical approach to some basic mine evaluation problems on the Witwatersrand. *Johannesburg Chemistry Metallurgy Mining Society South Africa*, v.52, p.151-163, 1951.
- Lima, J. S. S.; Silva, S. A.; Oliveira, R. B.; Cecílio, R. A.; Xavier, A. C. Variabilidade temporal da precipitação mensal em Alegre – ES. *Revista Ciência Agronômica*, v. 39, p. 327-332, 2008.
- Matheron, G. Principles of geostatistics. *Economic Geology*, v.58, p.1246-1266, 1963.
- Melo Junior, J. C. F.; Sedyama, G. C.; Ferreira, P. A.; Leal, B. G.; Minusi, R. B. Distribuição espacial da frequência de chuvas na região hidrográfica do Atlântico, Leste de Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.10, p.417-425, 2006.
- Mello, C. R.; Silva, A. M. da. Modelagem estatística da precipitação mensal e anual e no período seco para o estado de Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.13, p.68-74, 2009.
- Mingoti, S. A.; Neves, O. F. A metodologia de geoestatística aplicada na análise de séries temporais. *Revista Escola de Minas*, v.52, p.182-187, 1999.
- Mingoti, S. A.; Glória, F. A. A. *PrevGeo – A Geostatística aplicada em estudos de séries temporais: Uma comparação com o método de Holt-Winters e Box e Jankins*. *Revista Produto e Produção*, v. 6, p.24-32, 2002.
- Silva, M. I. S.; Guimarães, E. C.; Mavares, T. Previsão da temperatura média mensal de Uberlândia, MG, com modelos de séries temporais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.12, p.480-485, 2008.
- Webster, R.; Oliver, M. A. *Statistical methods in soil and land resource survey*. Oxford: Oxford University Press, 1990. 316p.
- Wei, W. W. S. *Time series analysis: Univariate and multivariate methods*. Philadelphia: Addison-Wesley, 1990. 496p.