

CORRELAÇÃO CRUZADA E MODELO DE REGRESSÃO NÃO LINEAR LOGÍSTICO NA DETERMINAÇÃO DAS RELAÇÕES HIDROLÓGICAS DO PANTANAL

Raphael Antonio do Prado Dias¹, Carlos Tadeu dos Santos Dias²,
Carlos Roberto Padovani³, Carlos Alberto Vettorazzi⁴

Resumo: *O regime de inundações e secas no Pantanal determina a distribuição e a abundância da fauna, da vegetação e dos processos ecológicos. A pecuária e a pesca, que são as principais atividades econômicas nesta planície de inundação, são extremamente influenciadas e dependentes desse regime. Os dados utilizados correspondem a valores quinzenais do nível do rio Paraguai em Ladário e da área inundada no Pantanal para o período de 2000 a 2009. Este estudo teve por objetivo definir, a partir da correlação cruzada, a defasagem entre as séries temporais das variáveis mencionadas anteriormente e ajustar o modelo não linear logístico a fim de estabelecer a relação entre estas.*

Palavras-chave: Modelo não linear, modelo logístico, correlação cruzada, método de Gauss-Newton.

Introdução

No Pantanal, a pecuária extensiva e a pesca são as principais atividades econômicas, conduzidas em estreita relação com o ambiente natural e, portanto, extremamente dependentes dele. A sazonalidade das inundações condiciona a produção de peixes e de demais organismos aquáticos, além de influenciar a produção pecuária e a distribuição e a abundância da vegetação e da fauna terrestre.

Apesar da importância do conhecimento da dinâmica da inundação para o Pantanal, esse processo ainda não é bem compreendido. Isso se deve, principalmente, à falta do emprego de métodos de análise e tecnologias que permitam caracterizar e monitorar de forma contínua o ambiente, além de permitir que se façam cenários e previsões do comportamento do sistema como um todo.

Foi utilizada a função de correlação cruzada para medir o grau de associação entre as séries temporais do nível do rio Paraguai em Ladário e da área inundada no Pantanal, considerando vários períodos de defasagem.

A relação entre o nível do rio Paraguai em Ladário e a área inundada no Pantanal apresenta formato sigmóide, a qual pode ser ajustada por modelos lineares e não lineares. Uma das vantagens de se utilizar modelos não lineares, em detrimento dos modelos lineares, reside na interpretação dos parâmetros, além de geralmente fornecer um bom ajuste com menos parâmetros, embora as estimativas destes sejam mais difíceis de se obterem.

¹IFSULMEMINAS - Campus Muzambinho,
raphael.dias@eafmuz.gov.br

²LCE, ESALQ/USP

³EMBRAPA - Pantanal

⁴LEB, ESALQ/USP

Objetivo

Este trabalho teve como objetivo obter a defasagem entre as séries temporais de nível do rio Paraguai em Ladário e a área inundada no Pantanal, a partir da função de correlação cruzada, e estimar os parâmetros do modelo não linear logístico pelo método de convergência de Gauss-Newton, a fim de estabelecer o relacionamento entre estas variáveis e, conseqüentemente, avaliar as relações hidrológicas do Pantanal.

Metodologia

Os dados deste trabalho foram obtidos na bacia do alto do rio Paraguai, entre as latitudes 14°S e 23°S e as longitudes 53°W e 61°W no Pantanal, envolvendo os países Brasil, Bolívia e Paraguai (Figura 1), no período de 2000 a 2009, em medidas quinzenais, totalizando 224 observações.

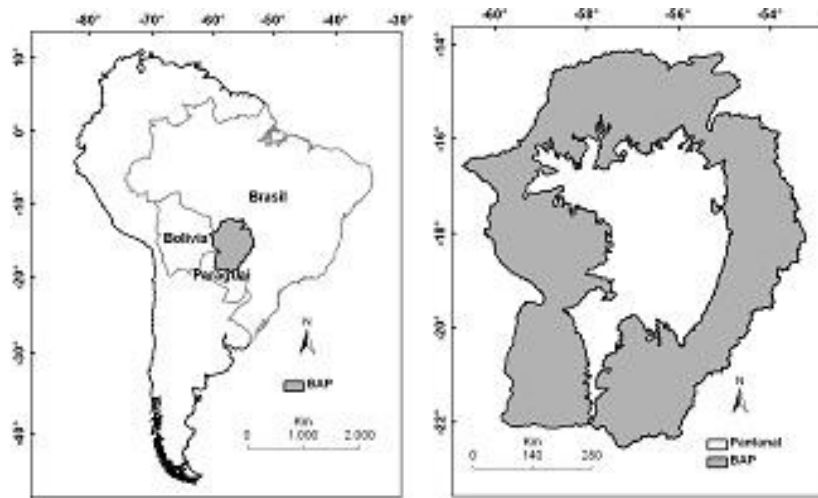


Figura 1: Localização da bacia do alto rio Paraguai no Brasil, Bolívia e Paraguai (América do Sul) e o limite do Pantanal dentro da bacia do alto Paraguai

A função correlação cruzada mede a correlação entre as séries temporais em diferentes períodos de tempos [7]. Ela foi utilizada para mensurar o grau de associação entre o nível do rio Paraguai em Ladário e a área inundada no Pantanal. De acordo com [1], o estimador da função de correlação cruzada é dado pela fórmula:

$$\hat{\rho}_{xy}(h) = \frac{\sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y})}{n^{-1} \sum_{t=1}^n (x_t - \bar{x})^2 \sum_{t=1}^n (y_t - \bar{y})^2}$$

em que x_t e y_t são as séries temporais; \bar{x} e \bar{y} são as médias; h é o coeficiente de defasagem entre as séries; e n é o número de observações.

O modelo de regressão não linear utilizado foi o logístico ([4], [2]) descrito pela equação:

$$Y_i = \frac{\alpha}{1 + \beta e^{-kx}} + \varepsilon_i$$

em que α representa a assíntota horizontal quando $x \rightarrow \infty$; β representa o valor associado com

o ponto de inflexão; k é um indicador da velocidade com que a variável resposta se aproxima do valor assintótico e ε_i são os erros que tradicionalmente são supostos como variáveis aleatórias independentes e identicamente distribuídos com distribuição normal, média zero e variância σ^2 .

Os modelos não lineares são mais difíceis de especificar e estimar que os modelos lineares e, além disso, as soluções são determinadas iterativamente [5]. O método de convergência utilizado foi o de Gauss-Newton, desenvolvido por [3], também conhecido como método da linearização que usa uma expansão em série de Taylor para aproximar o modelo de regressão não linear com termos lineares e, então, aplicar mínimos quadrados ordinários para estimar os parâmetros. Iterações desses passos geralmente conduzem a uma solução para o problema de regressão não linear. As análises e resultados foram obtidos através dos procedimentos *proc arima* e *proc nlin* do programa computacional *SAS* (2004).

Resultados e Discussão

O valor do coeficiente de defasagem h que maximizou a correlação cruzada entre as séries foi $h = -3$ (Figura 2), com $\hat{\rho}_{xy}(-3) = 0,91165$, indicando que a inundação na bacia do Pantanal influencia no nível do rio Paraguai em Ladário após 45 dias.

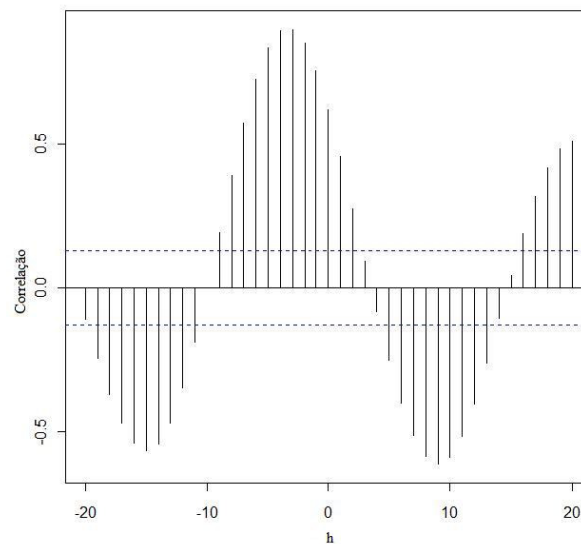


Figura 2: Correlação entre as séries em função da defasagem (h), obtidos a partir da função de correlação cruzada

Retirando o efeito da defasagem entre as séries, ajustou-se o modelo não linear logístico para a variável dependente nível do rio em Ladário em função da área inundada no Pantanal, sendo este significativo ao nível 0,05 de probabilidade ($p < 0,0001$), obtendo-se as estimativas dos parâmetros com seus respectivos erros padrão aproximado:

- $\alpha = 495,90(10,89)$
- $\beta = 6,41(0,51)$
- $k = 0.000129(7,44 \times 10^{-6})$

A partir das estimativas obtidas pelo método de Gauss-Newton, com 5 interações e critério de convergência 10^{-9} , ajustou-se a curva aos valores observados e estes se encontram na Figura 3.

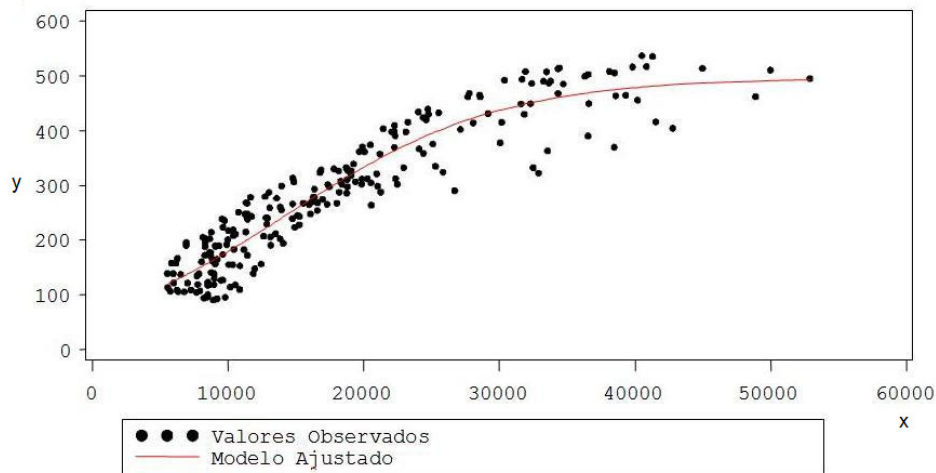


Figura 3: Curva do modelo logístico ajustado aos dados de nível do rio Paraguai em Ladário (y cm) em função da área inundada no Pantanal (x km²)

Conclusões

A estimativa da defasagem entre as séries temporais é de 45 dias, indicando que o nível do rio Paraguai em Ladário é influenciado 45 dias após pela área inundada no Pantanal e o modelo logístico se ajustou significativamente aos dados, indicando um nível do rio Paraguai assintótico de 495,9 cm e ponto de inflexão quando $x = 14402$ km², cuja velocidade máxima é de 0,016 cm/km².

Referências

- [1] CHATFIELD, C., *The analysis of time series: An introduction*. 5. ed. Chapman and Hall/CRC, 2004.
- [2] DRAPER, N.R; SMITH, H., *Applied regression analysis*. 3 ed. Wiley-Interscience, 1998.
- [3] HARTLEY, H.O., The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. *Technometrics*, **3**, 269-280, 1961.
- [4] LAIRD, A.K; HOWARD, A., Growth curves in inbred mice. *Nature*, **213(5078)**, 786-788, 1967.
- [5] RATKOWSKY, D.A., *Nonlinear regression modelling*. New York and Basel: Marcel Dekker, 1983.
- [6] SAS INSTITUTE., *Statistical analysis system user's guide: version 9.1*. Cary. 2004.
- [7] VANDAELE, W., *Applied time series and box-jenkins models*. Academic Press, 1983.

Agradecimentos

Os autores agradecem o apoio recebido da ESALQ pela oportunidade da realização desse trabalho e pelo apoio financeiro da CAPES e do IFSULDEMINAS.