

# APLICAÇÃO DA REGRESSÃO LINEAR MÚLTIPLA PARA A PREVISÃO MENSAL DE PRECIPITAÇÃO EM MACEIÓ-ALAGOAS

Washington L. F. Correia Filho<sup>1</sup>, Alexandre S. dos Santos,<sup>1</sup> Paulo S. Lucio<sup>2</sup>

<sup>1</sup> Meteorologista, Doutorando do PPG em Ciências Climáticas, UFRN, Brasil, Natal-RN,  
E-mails: washington@dca.ufcg.edu.br; puff78@gmail.com

<sup>2</sup> Professor Adjunto do Departamento de Estatística, UFRN, Brasil, Natal-RN, E-mail: pslucio@ccet.ufrn.br

Apresentado no XVII Congresso Brasileiro de Agrometeorologia – 18 a 21 de Julho de 2011 -  
SESC Centro de Turismo de Guarapari, Guarapari - ES

**RESUMO:** A utilização de métodos estatísticos na área de ciências climáticas tem contribuído para prever com maior confiabilidade as diversas variáveis meteorológicas, tais como a temperatura à superfície do mar (TSM), interligada ao regime pluviométrico. Neste trabalho foram utilizados dados mensais de TSM das 3 regiões do Atlântico (Norte, Sul e Tropical) e 4 regiões do Pacífico (Niño 1.2, Niño 3, Niño 4 e Niño 3.4) e de precipitação oriundos da cidade de Maceió para o período de 1961 a 2010, foram aplicados a Função de Correlação Cruzada (FCC) e Fator de Inflação da Variância (FIV) para identificação de lags da TSM, com o objetivo de prever a precipitação para o mês de julho de 2010. Constatamos que o resultado de p-valor significativo com o valor de 6%, porém com baixo valor do  $R^2$  de 30%. Os valores de VIF com exceção do lag<sub>5</sub> da região Niño 3.4 foram inferiores a 10, detectando multicolinearidade entre as séries, e a precipitação prevista para o mês de julho foi de 160,25 mm.

**PALAVRA-CHAVE:** Regressão de ridge, Temperatura à Superfície do Mar, Fator de Inflação da Variância.

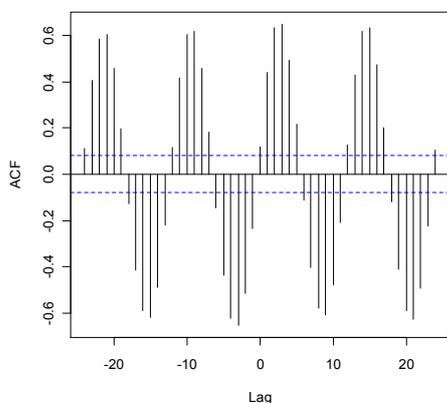
## MULTIPLE LINEAR REGRESSION FOR MONTHLY PRECIPITATION PREDICTION IN MACEIÓ-ALAGOAS

**ABSTRACT:** The use statistical methods in the climate sciences has contributed to more reliably predict the various meteorological variables such as sea surface temperature (SST), closely associated with rainfall. In this study, we employ SST monthly data of the three Atlantic regions (North, South and Tropical) and the four Pacific regions (Niño 1.2, Niño 2, Niño 4 and Niño 3.4) and precipitation from Maceió City for the period 1961 to 2010, were applied to Cross-Correlation Function (CCF) and the Variance Inflation Factor (VIF) to identify the SST lags with objective order to predict rainfall for the month of July 2010. We note that the result of significant the p-value of 6%, but with low value of 30% for  $R^2$ . The VIF values except for the SST lag<sub>5</sub> of the Niño 3.4 region were lower than 10, detecting multicollinearity between sets, and rainfall predicted of July was 160.25 mm.

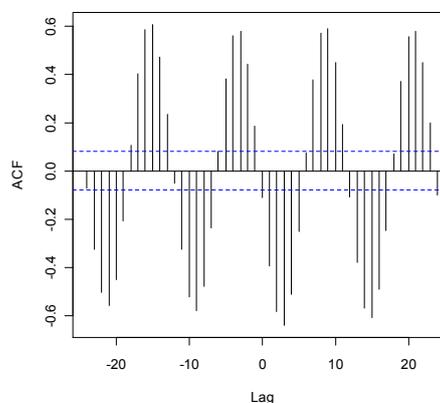
**KEYWORDS:** Ridge Regression, Sea Surface Temperature, Variance Inflation Factor.

**INTRODUÇÃO:** A aplicação de métodos estatísticos em ciências climáticas vem buscando identificar e prever com maior confiabilidade a temperatura à superfície do mar (TSM), visto que esta associada a outras variáveis meteorológicas detectam sinais que refletem diretamente no regime pluviométrico sobre as regiões do globo, e em particular sobre o Nordeste do Brasil. Cropper (1984) investigou o grau de correlação entre as variáveis climáticas mensais (multicolinearidade), da temperatura e precipitação, com ajustes de dados com alta qualidade para a região oeste dos EUA, utilizando dados mensais de 32 estações meteorológicas de com o objetivo de observar a análise multivariada. O autor concluiu que houve uma forte colinearidade para curto período de dados. Meisner (1979) aplicou a técnica de regressão de ridge para extrapolação de normais de precipitação de curtos registros utilizando 700 estações meteorológicas com vários parâmetros a dados mensais e anuais de precipitação no ano de 1975 para a ilha de Oahu, Hawai - EUA. O autor constatou método de regressão de ridge que estima com maior precisão os coeficientes de correlação reais, além de ser um estimador robusto de tendência local para extrapolação das normais de chuva do Hawai. Com base nestas informações, o objetivo deste trabalho será prever a precipitação através do modelo de regressão de ridge avaliando a variabilidade do oceano Atlântico e Pacífico, realizando a previsão para o mês de julho de 2010 para a estação oriunda na cidade de Maceió.

**MATERIAIS E MÉTODOS:** Para a realização deste trabalho, foram utilizados dados mensais de precipitação oriundas de uma estação meteorológica de Maceió, pertencente ao Instituto Nacional de Meteorologia (INMET) do período de janeiro de 1961 a junho de 2010, além de dados mensais de Temperaturas à Superfície do Mar (TSM) do National Oceanic and Atmospheric Administration (NOAA) das 3 regiões do Atlântico (Norte, Sul e Tropical) e 4 regiões do Pacífico (Niño 1.2, Niño 2, Niño 4 e Niño 3.4). Para a determinação da variável resposta, neste caso a precipitação para o mês de julho de 2010, utilizou-se o mecanismo da Função de Correlação Cruzada (FCC) ao qual confronta as séries de precipitação versus TSM, com o critério de identificar as defasagens para cada uma das regiões, bem como sua influência para prever o grau de associação entre elas. O resultado da FCC que pode ser visto na Figura 1, que mostra lags de 2, e 3 meses para as regiões norte (Fig. 1a) e tropical (Fig. 1b) do Atlântico, para a região sul (Fig. 1c) são os lags de 3 e 4 meses, já para as regiões do Pacífico foram observados lags de 2 e 3 meses para Niño 1.2 (Fig. 1d), para o Niño 3 foi identificado os lags de 3 e 4 meses (Fig. 1e), e os lags 4 e 5 para a região de Niño 4 (Fig. 1f) e os lags 5 e 6 para as regiões de Niños 3.4 (Fig. 1g).



(a) AL x ATLS



(b) AL x ATLN

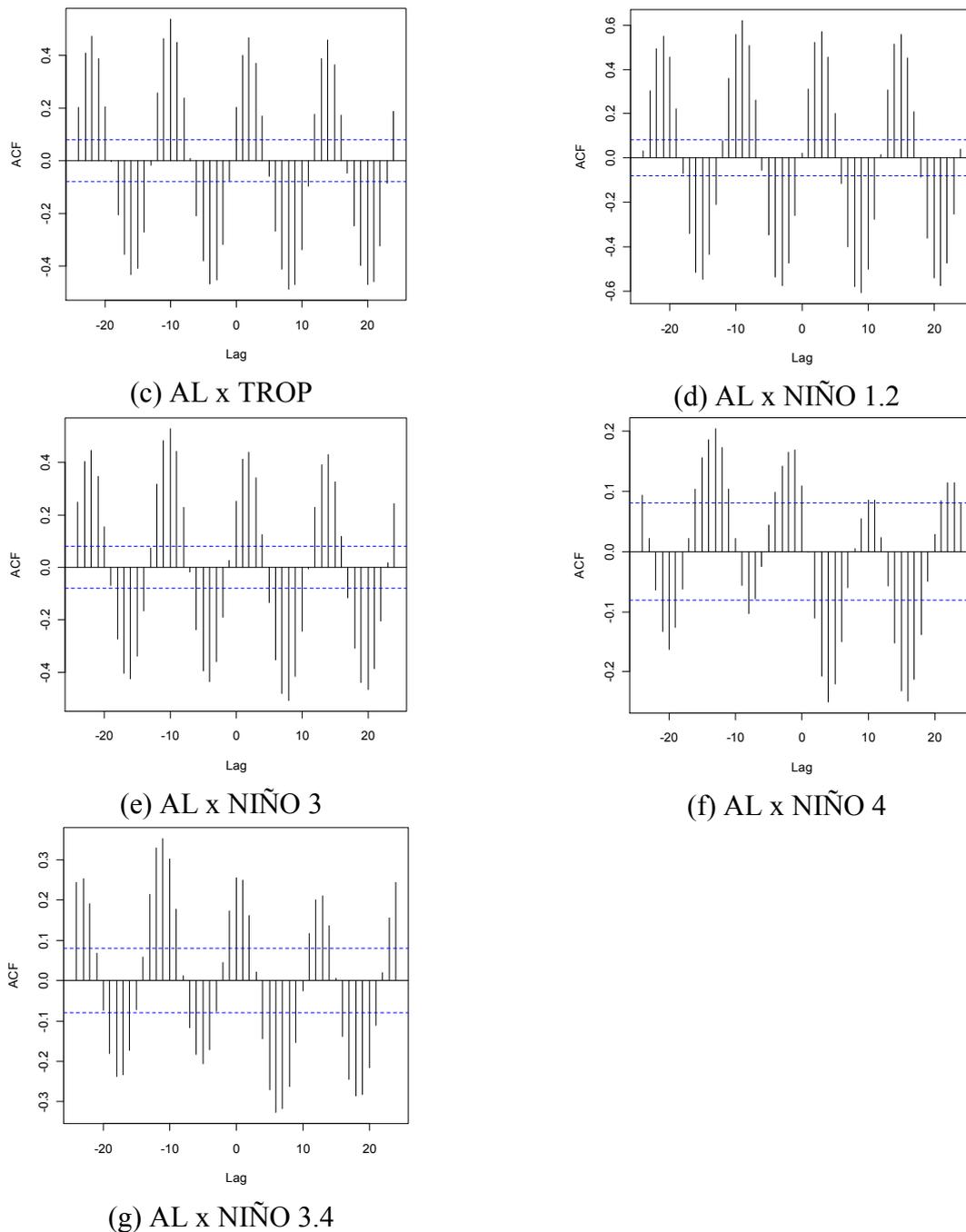


Figura 1 – Resultado das defasagens de FCC entre a precipitação e a TSM das regiões do Atlântico e Pacífico.

Para a detecção da presença de multicolinearidade foi utilizado o Fator de Inflação da Variância (FIV), em que estas medidas utilizam a variância  $n$ , estimada pelo coeficiente de regressão descritas pela equação 1 (KUTNER *et al.*, 2004). Para valores maiores do que 1, há indicação da presença do referido efeito, quando o valor obtido excede 10, o efeito é problemático e deve ser tratado. Na correção do efeito por meio do modelo de regressão de ridge, o usual estimador de mínimos quadrados (equação 2) é modificado, resultando no estimador viesado ridge (equação 3) (HIDALGO-CARNEIRO *et al.* 2009).

$$FIV_k = \frac{1}{1 - r_k^2} \quad (1)$$

$$\beta = (X'X)^{-1} X'Y \quad (2)$$

$$\beta = (X'X + kI)^{-1} X'Y \quad (3)$$

**RESULTADOS E DISCUSSÃO:** Na aplicação do modelo de regressão de ridge foi observada a relação com o método dos mínimos quadrados através do programa R (R development core team, 2008) e também análise dos resultados e gráficos, além da composição dos traços de “ridge” apresentado na Figura 2, determinando o seu valor ideal.

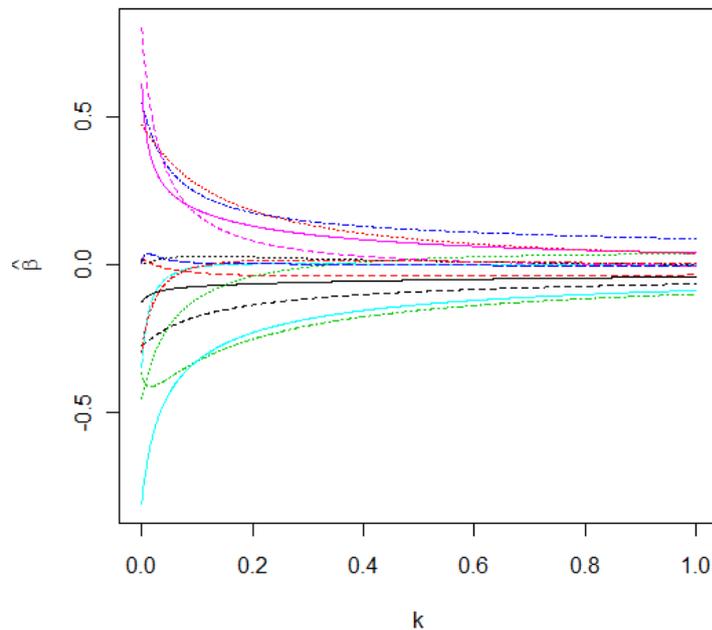


Figura 2 – Resultado dos traços da regressão de ridge.

Os resultados de VIF, conforme visto na estação de Maceió (Tabela 1) aponta que com exceção do lag<sub>5</sub> da região Niño 3.4, os demais os valores foram inferiores a 10, além disso, o teste p-valor é significativo, com valor de 6%. Outro fator a considerar é o pequeno grau de explicação da variabilidade do modelo (R<sup>2</sup>=0,30).

Tabela 1: Resultado dos valores de VIF e de suas respectivas regiões

Parâmetro	VIF	Parâmetro	VIF
NATL-lag2	6,0644	N1.2-lag3	4.7545
NATL-lag3	5.5724	N3-lag3	5.1144
SATL-lag2	5.6340	N3-lag4	8.3718
SATL-lag3	5.6205	N4-lag4	7.0998
TROP-lag3	7.6030	N4-lag5	8.3872
TROP-lag4	8.6765	N3.4-lag5	10.3875
N1.2-lag2	4.5096	N3.4-lag6	9.4631

A equação para a previsão de precipitação com a regressão de ridge é

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14}$$

$$\begin{aligned}
Y_{\text{jul}} = & 208,6047 - 30,36\text{NATL}_{\text{lag}2} + 1,72\text{NATL}_{\text{lag}3} - 110,52\text{SATL}_{\text{lag}2} + 148,14\text{SATL}_{\text{lag}3} \\
& - 65,07\text{TROP}_{\text{lag}3} + 161,59\text{TROP}_{\text{lag}4} - 25,68\text{N1.2}_{\text{lag}2} - 50,19\text{N1.2}_{\text{lag}3} - 88,41\text{N3}_{\text{lag}3} + 7,32\text{N3}_{\text{lag}4} \\
& - 126,74\text{N4}_{\text{lag}4} + 92,89\text{N4}_{\text{lag}5} + 1,21\text{N3.4}_{\text{lag}5} - 17,63\text{N3.4}_{\text{lag}6}
\end{aligned}$$

A precipitação prevista para o mês de julho de 2010 foi de 160,25 mm. O valor observado pela estação meteorológica de Maceió para o mês previsto foi de 218,4 mm. Logo, o modelo subestimou o valor da precipitação.

**CONCLUSÕES:** No presente verificou-se a necessidade da utilização de 14 lags para a composição do modelo de regressão de ridge, o valor de VIF das séries que constituíram os traços de ridge com exceção do lag<sub>5</sub> da região de Niño 3.4 foram inferiores a 10, indicando que houve multicolinearidade entre elas. Foi detectado valores de p-valor significativo com 6%, embora o R<sup>2</sup> tenha sido de de 30%. O valor previsto de precipitação pelo modelo de regressão de ridge foi de 160,25mm.

**AGRADECIMENTOS:** Paulo S. Lucio é Bolsista de Produtividade em Pesquisa (PQ2) do CNPq (Proc.302493/2007-7). Os dados das estações climatológicas foram gentilmente cedidos pelo INMET. O primeiro autor agradece à CAPES pela concessão da Bolsa de Doutorado.

**REFERÊNCIAS:** CROPPER, J.P. Multicollinearity within selected Western North American Temperature and Precipitation data sets. **Tree Bulletin**, Vol. 44, nº 28-37, 1984.

HIDALGO-CARNEIRO, L., SARMANHO, G. F., TORRES, F. C., SALVADOR, M. A., SILVA, F. D. S., FERREIRA, D. B., LUCIO, P. S., A Regressão Ridge como Modelo de Precipitação Sazonal, **In: XI Escola de Modelos de Regressão**, 2009, Recife. Anais da XI Escola de Modelos de Regressão. São Paulo: ABE, 2009. V. 1. p. 1-4.

KUTNER, M. H., NACHTSHEIM, C. J., NETER, J. LI, W. **Applied Linear Statistical Models**. 5ª edição. McGraw-Hill-Irwin. 2004.

MEISNER, B. N. Ridge Regression – Time Extrapolation Applied to Hawaiian Rainfall Normals, **J. of Appl. Meteor.** Vol.18, nº 904-912, 1979.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. Vienna, 2008. Disponível em: <<http://www.r-project.org>>, 2010