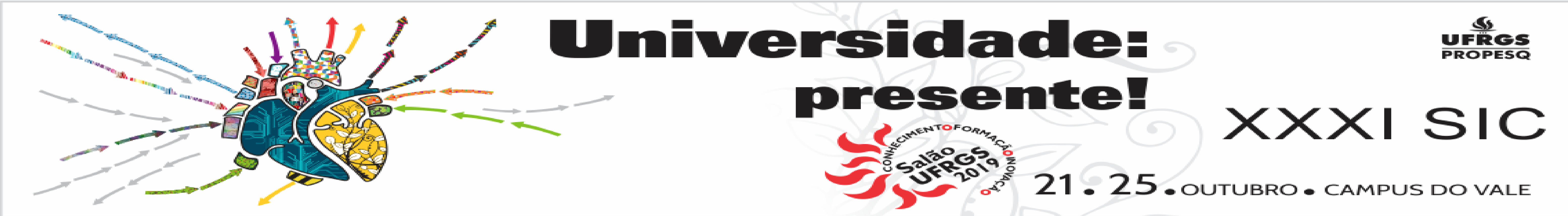

Análise Hierárquica para Dados de Bacias Hidrográficas

Verônica Stamatto Peres

Orientadora: Dra. Sílvia Regina Costa Lopes

IME - UFRGS - Brasil

1. Introducão

O objetivo deste projeto é analisar e desenvolver modelos lineares generalizados para séries temporais. Para isso utilizamos dois bancos de dados diferentes. Primeiramente, usamos os dados referentes à medida de qualidade da água em quatro pontos do Rio dos Sinos. O segundo banco de dados refere-se a vazão de água do Rio Colorado, Estados Unidos.

Analisamos os dados de forma descritiva e ralizamos simulações modelando a série temporal histórica para esses dados, aplicando a metodologia hierárquica Bayesiana. Com a análise descritiva e as simulações realizadas para o primeiro banco de dados conseguimos desenvolver outro modelo para as séries temporais, que está sendo testado com os dados do Rio Colorado.

Até o momento, esse modelo apresentou melhor estimativa para os parâmetros quando analisamos dados do Rio Colorado em relação aos modelos estudados anteriormente. Os resultados obtidos serão apresentados através de tabelas e gráficos.

2. Séries Temporais

O trabalho iniciou-se com uma análise descritiva da série temporal do Rio dos Sinos.

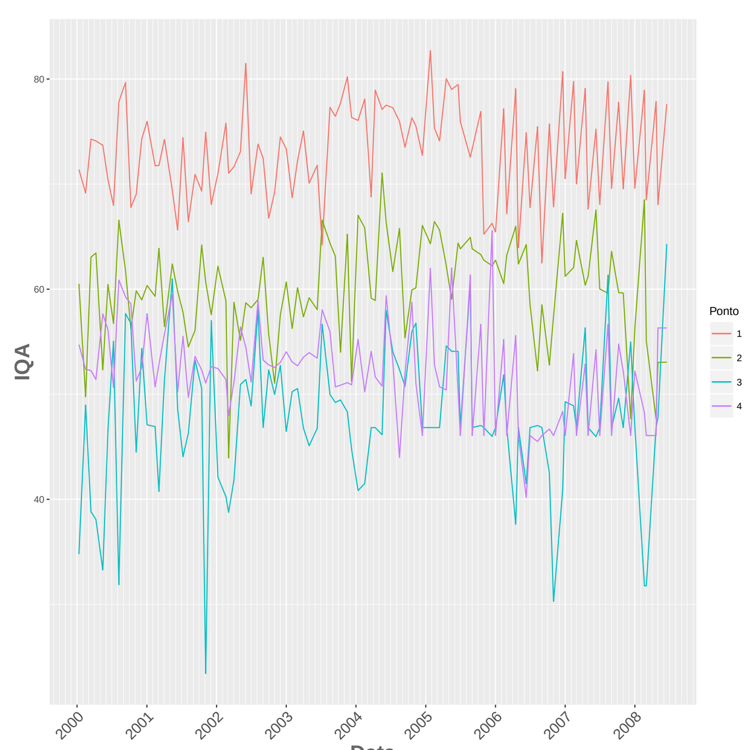

Figura 1: Série temporal histórica da qualidade da água do Rio dos Sinos, em quatro pontos de coleta.

Observamos a maior qualidade média da água no ponto de coleta 1 em relação aos demais, seguido pelo ponto de coleta 2.

A segunda série temporal se refere à dois pontos de coleta no Rio Colorado, analisando a vazão de água.

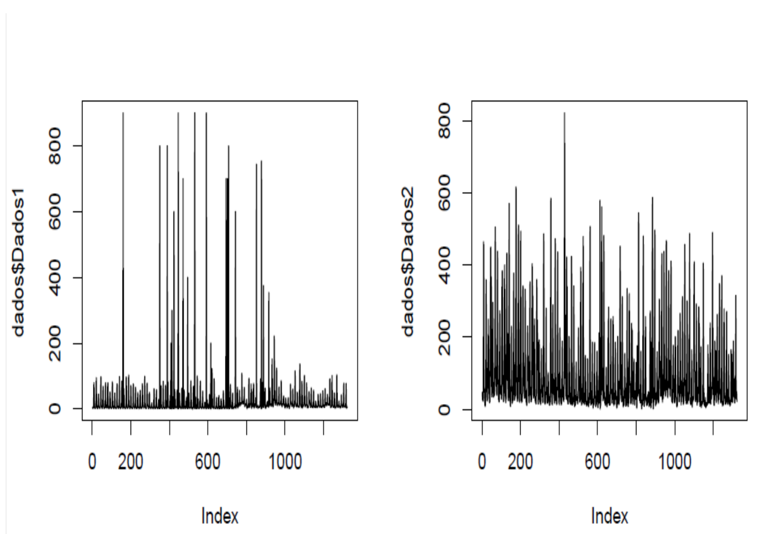

Figura 2: Série temporal histórica da vazão da água do Rio Colorado, em dois pontos de coleta.

3. Metodologia

A metodologia de interesse neste projeto consiste da análise da estrutura espaço-temporal nos modelos lineares generalizados, usando a metodologia Bayesiana. A estrutura temporal será modelada através de processos autoregressivos de primeira e segunda ordem.

O modelo de interesse neste trabalho é dado por

$$\mu_{it} \equiv E(Y_{it}) = \alpha_0 + S_i(t) + R_i,$$

onde Y_{it} é a contagem de um fenômeno aleatório, μ_{it} é sua esperança matemática, α_0 é uma constante, $S_i(t)$ é o efeito temporal associado a qualquer região i e R_i é o efeito aleatório associado a i-ésima região.

A componente R_i é a soma de efeitos aleatórios espacialmente correlacionados b_i e efeitos aleatórios não-correlacionados h_i, onde b_i será modelado através de processos auto-regressivos condicionais intrínsicos (CAR). Primeiramente incluíremos apenas os efeitos não-correlacionados h_i nas simulações.

Os dois modelos para explicar a componente de efeito temporal, $S_i(t)$ que consideramos neste trabalho são:

Tabela 1: Modelos utilizados para as séries temporais do Rio dos Sinos e Rio Colorado.

Modelo	$S_i(t)$	Especificação do Modelo
$M1$	$A_1(t)$	$A_1(t) \sim \mathcal{N}(\phi_t + \rho(A_1(t-1) - \phi_{t-1}), \sigma^2_{\epsilon_t})$
$M1^*$	$A_1(i,t)$	$A_1(i,t) \sim \mathcal{N}(\phi_t + \rho(A_1(i,t-1) - \phi_{t-1}), \sigma^2_{\epsilon_{i,t}})$
$M2$	$A_5(t)$	$A_5(t) \sim \mathcal{N}(\phi_t + \rho(\phi A_5(t-1))+$
		$\rho(A_5(t-12)) - \rho(\phi A_5(t-13)) - \rho(\phi_{t-1} - \phi_1))$
$M2^*$	$A_5(i,t)$	$A_5(i,t) \sim \mathcal{N}(\phi_t + \rho(\phi A_5(i,t-1))+$
		$\rho(A_5(i,t-12)) - \rho(\phi A_5(i,t-13)) - \rho(\phi_{t-1} - \phi_1))$

Os modelos $M1$, $M1^*$ possuem um processo auto-regressivo imposto na média da distribuição $apriori$ e apresentam um parâmetro ρ na sua estrutura auto-regressiva. Os modelos $M2$ e $M2^*$ são um auto-regressivo para uma série temporal com sazionalidade.

Na literatura, o parâmetro ρ é tal que $|\rho| < 1$, para garantir a convergência das cadeias de Markov quando do uso do algoritmo Monte Carlo Markov Chain.

A diferença entre os modelos Mj e Mj^*, onde $j \in \{2, 6\}$, é que para aqueles sem asterisco consideramos para todas as regiões, enquanto os modelos com asterisco são considerados diferentemente para cada i-ésima região.

4. Simulação

Serão apresentados os gráficos dos estimadores e médias das melhores simulações de cada série temporal, que se referem ao modelo $M2^*$ para os dados do Rio dos Sinos e o modelo $M6*$ para os dados do Rio Colorado.

Para comparar os modelos através de simulações, utilizaremos alguns critérios de seleção de modelo propostos em Figueira et al (2016). Os critérios selecionados foram: *Deviance Information Criterion* (DIC), *Logarithm of the Pseudo Marginal Likelihood* (LPML), *Pearson Residuals* (Qp) e *Widely Applicable Information Criterion* (WAIC) [ver Paulino et al.(2015)].

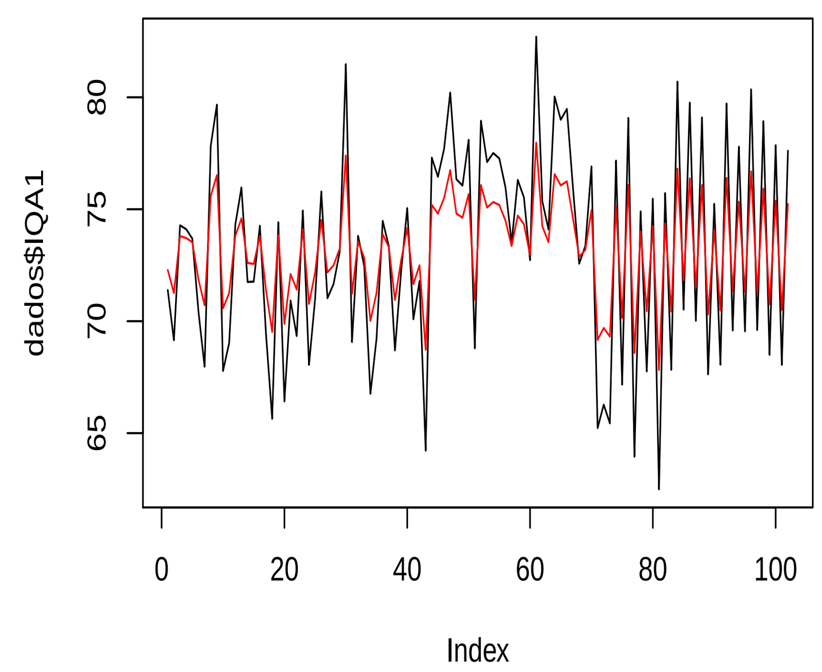

Figura 3: Gráfico de comparação entre os estimadores e os dados do Rio dos Sinos da região 1 para o modelo $M2*$.

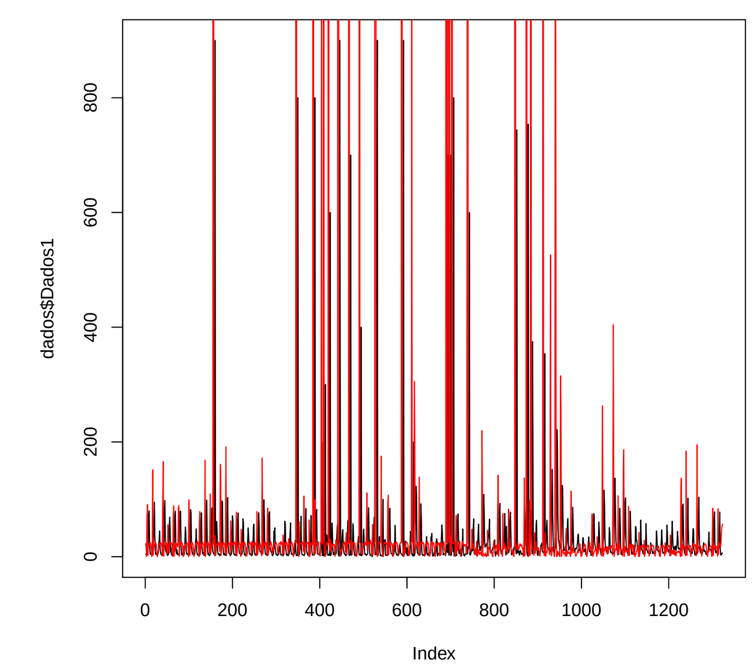

Figura 4: Gráfico de comparação entre os estimadores e os dados do Rio Colorado da região 1 para o modelo $M6*$.

5. Conclusão

Concluimos que o melhor modelo que se ajusta à série temporal do Rio dos Sinos é o $M1*$. Os modelos para a série temporal do Rio Colorado ainda estão sendo ajustados para obtermos melhores resultados. Até o momento, os gráficos das séries ajustadas apresentam picos extremos(tanto altos como baixos) que precisam ser analisados.

Referências

[1] Congdon, P. (2006). *Bayesian statistical modelling*. London: Wiley.

[2] Figueira, C.V., S.R.C. Lopes e G.L. Silva (2016). "Spatiotemporal Analysis by Autoregressive Models". Em preparação.

[3] Paulino, C.D., A.A. Turkman e B. Murteira (2009). *Estatística Bayesiana*. Lisboa: Fundação Calouste Gulbenkian.

[4] Rampaso, R.C., A.D.P. Souza e E.F. Flores (2015). "Bayesian analysis of spatial data using different variance and neighbourhood structures". *Journal of Statistical Computation and Simulation, DOI:10.1080 00949655.2015.1022549*.

vpstamatto@gmail.com
silvia.lopes@ufrgs.br