

Introdução

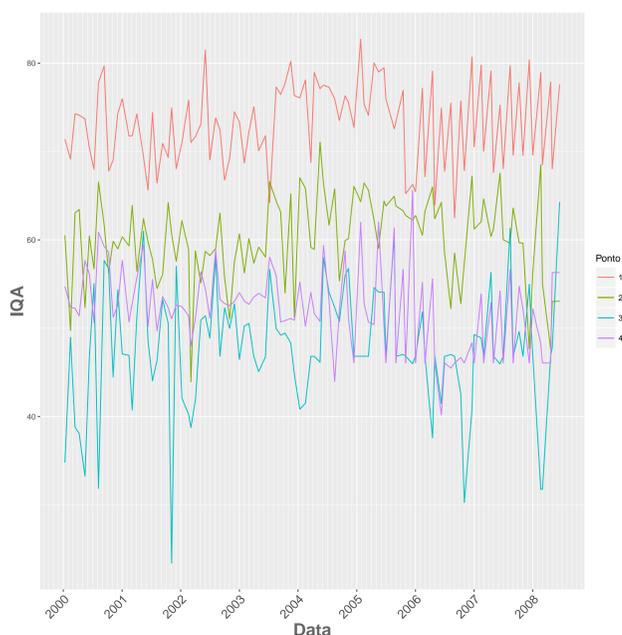
Este trabalho tem como objetivo principal modelar séries temporais históricas que compreendem um período de 9 anos. Estas séries temporais consistem de dados relativos à qualidade da água do Rio dos Sinos, cedidos pela FEPAM e coletados mensalmente em 4 pontos distintos do Rio dos Sinos durante o período de 11/01/2000 a 16/06/2008, resultando em 102 observações em cada ponto de coleta.

Normalmente, a qualidade da água é medida através da quantidade de coliformes (em mg/litro), pela quantidade de sólidos totais, pela turbidez, pela quantidade de sulfetos ou ainda pelo *iqa* (índice de qualidade da água). Neste trabalho utilizaremos o *iqa* como forma de mensurar a qualidade da água.

O interesse está na utilização do ambiente R com o *software* *OpenBUGS* para realizar simulações baseadas no método de Monte Carlo através de cadeias de Markov (MCMC).

Análise Descritiva das Séries Temporais

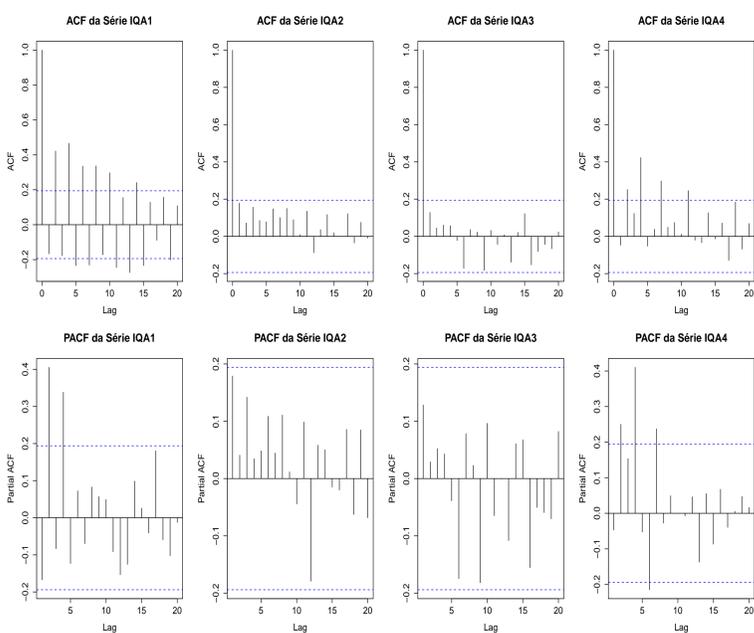
O trabalho iniciou-se com uma análise descritiva das séries temporais referentes aos quatro pontos de coleta ao longo do Rio dos Sinos:



Séries Temporais dos IQA's nos 4 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. O que resulta nesta degradação da qualidade da água no decorrer do curso do Rio dos Sinos?

Antes de testarmos como as séries temporais se ajustam aos modelos propostos, observamos os gráficos das funções de Autocorrelação (*acf*) e de Autocorrelação Parcial (*pacf*) para procurarmos algum padrão nas séries.



Gráficos das Funções de *acf* e *pacf* das Séries Temporais de IQA dos 4 Pontos de Coleta.

Identificamos as séries temporais de IQA dos pontos de coleta 2 e 3 como ruídos brancos, ou seja, não apresentam autocorrelação significativa. Os gráficos indicam, também, que as séries temporais dos pontos de coleta 1 e 4 aparentam ser processos Auto Regressivos.

Metodologia

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

O modelo de interesse neste trabalho é dado por

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

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, para $i \in \{1, \dots, 4\}$ e $t \in \{1, \dots, 102\}$ com $g(\cdot)$ uma função que relaciona as componentes espaço-temporal e aleatória.

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 intrínsecos (CAR). Primeiramente incluiremos apenas os efeitos não-correlacionados (h_i) nas simulações.

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

Modelo	$S_i(t)$	Especificação do Modelo
M1	δt	tendência linear aditiva para todas as 4 regiões
M1*	$\delta_i t$	tendência linear aditiva para a i -ésima região
M2	$A_1(t)$	$A_1(t) \sim \mathcal{N}(\phi_t + \rho(A_1(t-1) - \phi_{t-1}), \sigma_{\epsilon_t}^2)$
M2*	$A_1(i, t)$	$A_1(i, t) \sim \mathcal{N}(\phi_t + \rho(A_1(i, t-1) - \phi_{t-1}), \sigma_{\epsilon_{i,t}}^2)$
M3	$A_2(t)$	$A_2(t) \sim \mathcal{N}(\phi_1 A_2(t-1), \sigma_{\epsilon_t}^2)$
M3*	$A_2(i, t)$	$A_2(i, t) \sim \mathcal{N}(\phi_1 A_2(i, t-1), \sigma_{\epsilon_{i,t}}^2)$
M4	$A_3(t)$	$A_3(t) = \phi_1 A_3(t-1) + \epsilon_t$
M4*	$A_3(i, t)$	$A_3(i, t) = \phi_1 A_3(i, t-1) + \epsilon_{i,t}$
M5	$A_4(t)$	$A_4(t) = \phi_t + \rho(A_4(t-1) - \phi_{t-1}) + \epsilon_t$
M5*	$A_4(i, t)$	$A_4(i, t) = \phi_t + \rho(A_4(i, t-1) - \phi_{t-1}) + \epsilon_{i,t}$

Onde M1 e M1* são modelos que consideram apenas uma simples tendência linear aditiva ao efeito temporal. M2, M2*, M3 e M3* possuem um processo auto-regressivo imposto na média da distribuição *a priori*, porém, os modelos M2 e M2* apresentam um parâmetro ρ extra na sua estrutura auto-regressiva. Já os modelos M4, M4*, M5 e M5* consideram o processo auto-regressivo imposto aditivamente na componente $S_i(t)$, mas apenas os modelos M4 e M4* apresentam um parâmetro ρ extra nesta estrutura auto-regressiva aditiva.

O parâmetro ρ foi considerado por Waller et al. (1977b) como livre de qualquer imposição. Neste trabalho, consideramos $|\rho| < 1$ para garantir a convergência das cadeias de Markov quando do uso do algoritmo Monte Carlo Markov Chain.

A diferença entre os modelos M_j e M_j^* , onde $j \in \{1, 2, 3, 4, 5\}$, é que para os modelos sem asterisco os modelos são considerados para todas as 4 regiões, enquanto os modelos com asterisco são considerados diferentemente para cada i -ésima região, onde $i \in \{1, 2, 3, 4\}$.

Resultados e Conclusão

Serão apresentados os resultados das simulações realizadas até o presente momento, as quais incluem os modelos M1, M1*, M2 e M2* sem efeitos aleatórios espacialmente correlacionados (b_i), incluindo na componente R_i apenas os efeitos não-correlacionados (h_i).

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) [ver Spiegelhalter et al (2002)], *Logarithm of the Pseudo Marginal Likelihood* (LPML) [ver Souza (2015) e Congdon (2006)], *Pearson Residuals* (Qp) [ver Paulino et al (2009)] e *Widely Applicable Information Criterion* (WAIC) [ver Vehtari e Gelman (2015), Gelman et al (2014) e Watanabe (2013a, b)].

Modelo	Qp	LPML	DIC	WAIC1	WAIC2	σ_h^2
M1h	220.82	-521.55	2632.00	1043.05	1043.11	166.8196
M1h*	217.79	-520.72	2632.00	1041.24	1041.44	152.0680
M2h	230.01	-534.42	2670.00	1065.67	1068.60	5.07881
M2h*	258.81	-555.64	2709.00	1103.96	1111.40	9.3962

Comparação do ajuste dos modelos Mj/h, onde $j \in \{1, \dots, 5\}$.

Tendo em mãos os resultados obtidos até o presente momento, podemos afirmar que estamos nos direcionando corretamente para a escolha de um modelo que melhor se ajusta às séries temporais de IQA dos 4 pontos de coleta do Rio dos Sinos. Visto que a nossa proposta é finalizar as simulações para determinar qual o melhor modelo que descreve os dados apresentados, entre os modelos aqui estudados.

Bibliografia

- Congdon, P. (2006). *Bayesian statistical modelling*. London: Wiley.
- Figueira, C.V., S.R.C. Lopes e G.L. Silva (2016). *Spatiotemporal Analysis by Autoregressive Models*.
- Gelman, A., J. Hwang e A. Vehtari (2014). *Understanding predictive information criteria for Bayesian models*. Statistics and Computing, vol. 24(6), 997-1016.
- Paulino, C.D., A.A. Turkman e B. Murteira (2009). *Estatística Bayesiana*. Lisboa: Fundação Calouste Gulbenkian.
- 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.
- Spiegelhalter, D.J., N.G. Bestm, B.P. Carlin, e A. Van der Linde (2002). *Bayesian Measures of Model Complexity and Fit (with Discussion)*. Journal of the Royal Statistical Society, Series B, Vol. 64(4),583-616.
- Vehtari, A. e A. Gelman (2015). *WAIC and cross-validation in Stan*. Disponível em http://www.stat.columbia.edu/~gelman/research/unpublished/waic_stan.pdf. Acessado em 2 de Agosto de 2016.
- Watanabe, S. (2013a). *WAIC and WBIC are information criteria for singular statistical model evaluation*. Disponível em http://www.me.inf.kyushu-u.ac.jp/witmse2013/abst_files/watanabe2013.pdf. Acessado em 2 de Agosto de 2016.
- Watanabe, S. (2013b). *A widely applicable Bayesian information criterion*. Journal of Machine Learning Research, vol. 14(3), 867-897.

maxdenig@hotmail.com
silvia.lopes@ufrgs.br