

Elastic net combinado com ondaletas para análise de agrupamentos de dados de níveis do mar

Ana Cláudia Festucci de Herval ¹, Thelma Sáfadi ², Leila Maria Ferreira ³

Introdução

A compreensão à respeito do comportamento do nível do mar é de extrema importância, uma vez que seu aumento ou decaimento poderá causar diversos prejuízos para toda a população, sendo eles ambientais e/ou socioeconômicos, dentre outros fatores. A literatura mostra que a média global do nível do mar aumentou algo em torno de 1.75 e 1.91 mm/ano na última metade do século XX (Solomon et al., 2007 e Church et al., 2004). Segundo Dasgupta et al. (2009), o aumento do nível do mar nos países em desenvolvimento deslocarão dezenas de milhões de pessoas no mundo no século atual.

Conceitualmente, a transformada de ondaleta é uma técnica para visualizar e representar um sinal, que será decomposto em níveis de resolução, de forma que cada nível adiciona detalhes. Matematicamente, é representado por uma função que oscila no tempo ou no espaço. O método tem janelas deslizantes que expandem ou comprimem para capturar sinais de baixa e alta frequência, respectivamente (Percival e Walden, 2000). Ferreira et al. (2018) justifica o uso de ondaletas por ser uma ferramenta que permite a extração de características que estão ocultas, aumentando assim a precisão dos resultados.

Zou e Hastie (2015) propuseram o elastic net como um novo método de regularização e seleção de variável que também incentiva um efeito de agrupamento, onde preditores fortemente correlacionados tendem a estar dentro ou fora do modelo juntos. O elastic net é um método de regressão regular que combina linearmente as penalidades L_1 e L_2 dos métodos de regressão LASSO e Ridge. Sua principal característica é o agrupamento de variáveis correlacionadas, onde o número de preditores é maior que o número de observações. Sáfadi (2017) mostrou que a técnica de combinar elastic net e transformada de ondaletas foi eficaz para análise de agrupamentos de séries temporais.

Este trabalho pretende utilizar a técnica de elastic net combinado com a decomposição por ondaletas com o objetivo de fazer uma análise de agrupamentos de diferentes séries de nível do mar em níveis de detalhes que a série original não fornece. Foi utilizado o software R para a realização das análises presentes (R Core Team, 2017).

Material e Método

Foram utilizados dados mensais de nível do mar, correspondentes ao período de Janeiro de 1986 à Dezembro de 2011, obtidos em diferentes estações de coleta conforme apresentadas na Tabela 1. Os dados foram obtidos de *Global Sea Level Observing System* (GLOSS), através de *Fast mode data delivery*, cujos dados são distribuídos como valores por hora, diários e mensais.

¹Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Departamento de Estatística, Universidade Federal de Lavras (UFLA). e-mail: anafestucci@outlook.com

²Departamento de Estatística, Universidade Federal de Lavras (UFLA). e-mail: safadi@ufla.br

³Departamento de Estatística, Universidade Federal de Lavras (UFLA). e-mail: leilamaria2003@yahoo.com.br

Tabela 1: Localização das estações utilizadas para coleta das séries de nível do mar.

Estação	Latitude	Longitude	Localização
ABASHIRI	44 01N	144 17E	Japan
BRISBANE	27 22S	153 10E	Australia
ILHA FISCAL	22 54S	043 10W	Brazil
KODIAK ISLAND	57 44N	152 31W	USA Alaska
LANGKAWI	06 26N	099 45E	Malaysia
QUARRY BAY	22 18N	114 13E	Hong Kong, PRC
TANJONG PAGAR	01 16N	103 51E	Singapore
KO TAPHAO NOI	07 50N	098 26E	Thailand
VARDO	70 20N	031 06E	Norway
WELLINGTON	41 17S	174 47E	New Zealand

Fonte: Sáfadi(2014).

A Figura 1 mostra a localização dessas dez estações no planeta e permite observar a proximidade das mesmas.

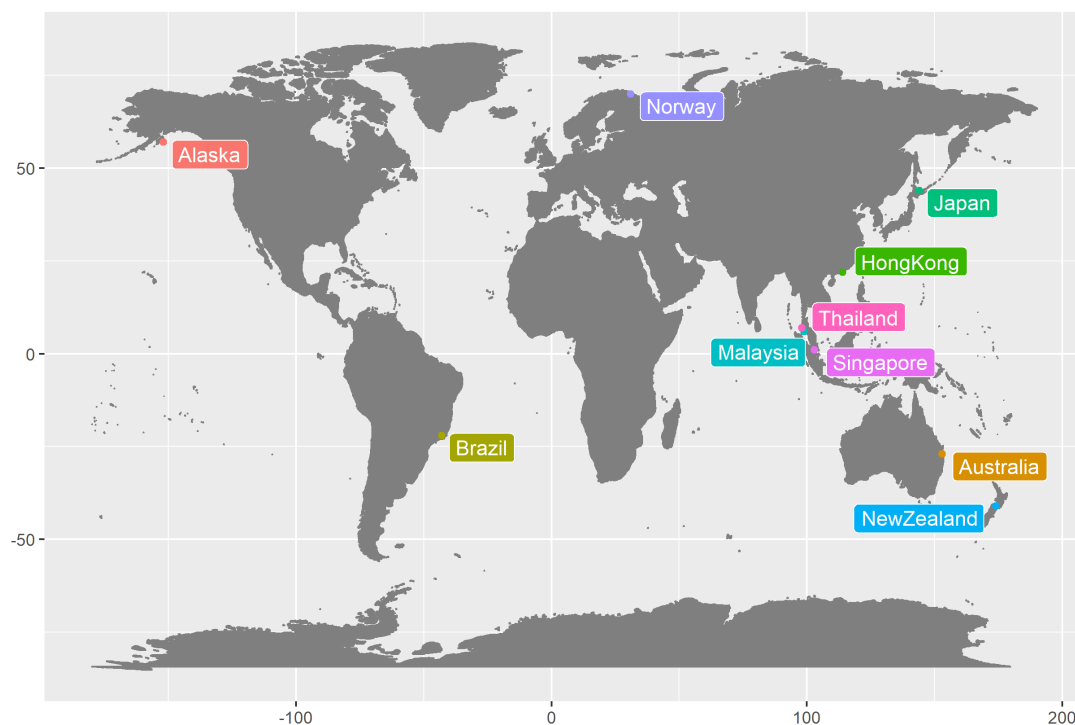


Figura 1: Localização no mapa do mundo das estações estudadas.

Fonte: do autor.

Cada uma das séries foram decompostas utilizando-se a transformada discreta de ondas não decimada (conhecida em inglês por *non-decimated wavelet transform* - NDWT). Utilizamos a ondaleta de Daubechies com 8 momentos nulos e 4 níveis de decomposição. Como resultado, obtemos 4 séries temporais compostas pelos coeficientes de detalhes dos níveis de decomposição (d1, d2, d3 e d4) obtidos pela transformada.

Para prosseguir com a análise de agrupamento, o elastic net foi aplicado a cada um dos níveis de detalhe de cada uma das dez séries.

Resultados

A Figura 2 apresenta as séries padronizadas pela média e desvio padrão do nível do mar para os diferentes países. Inicialmente, pode-se observar que Norway, Japan, Singapore, Honk Kong e Alaska apresentam um nível do mar crescente mais evidente que os demais. Thailand e Malaysia apresentam um crescimento inicial para então estabilizar. Por último, Brazil, Australia e New Zealand apresentam um comportamento mais constante ao longo do tempo.

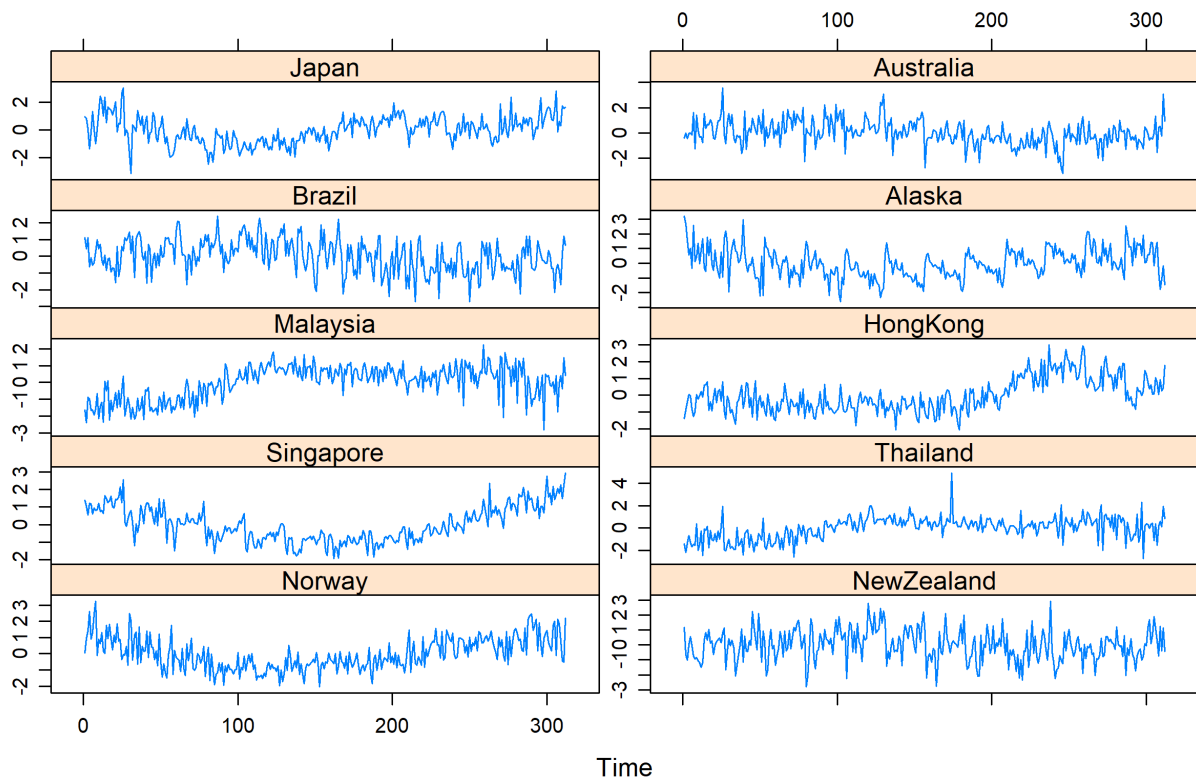


Figura 2: Séries de nível do mar no período de Jan/1986 à Dez/2011.

Fonte: do autor.

Os agrupamentos obtidos pelo elastic net para cada nível de detalhe estão apresentados na Figura 3 e o resumo com os grupos encontram-se na Tabela 2.

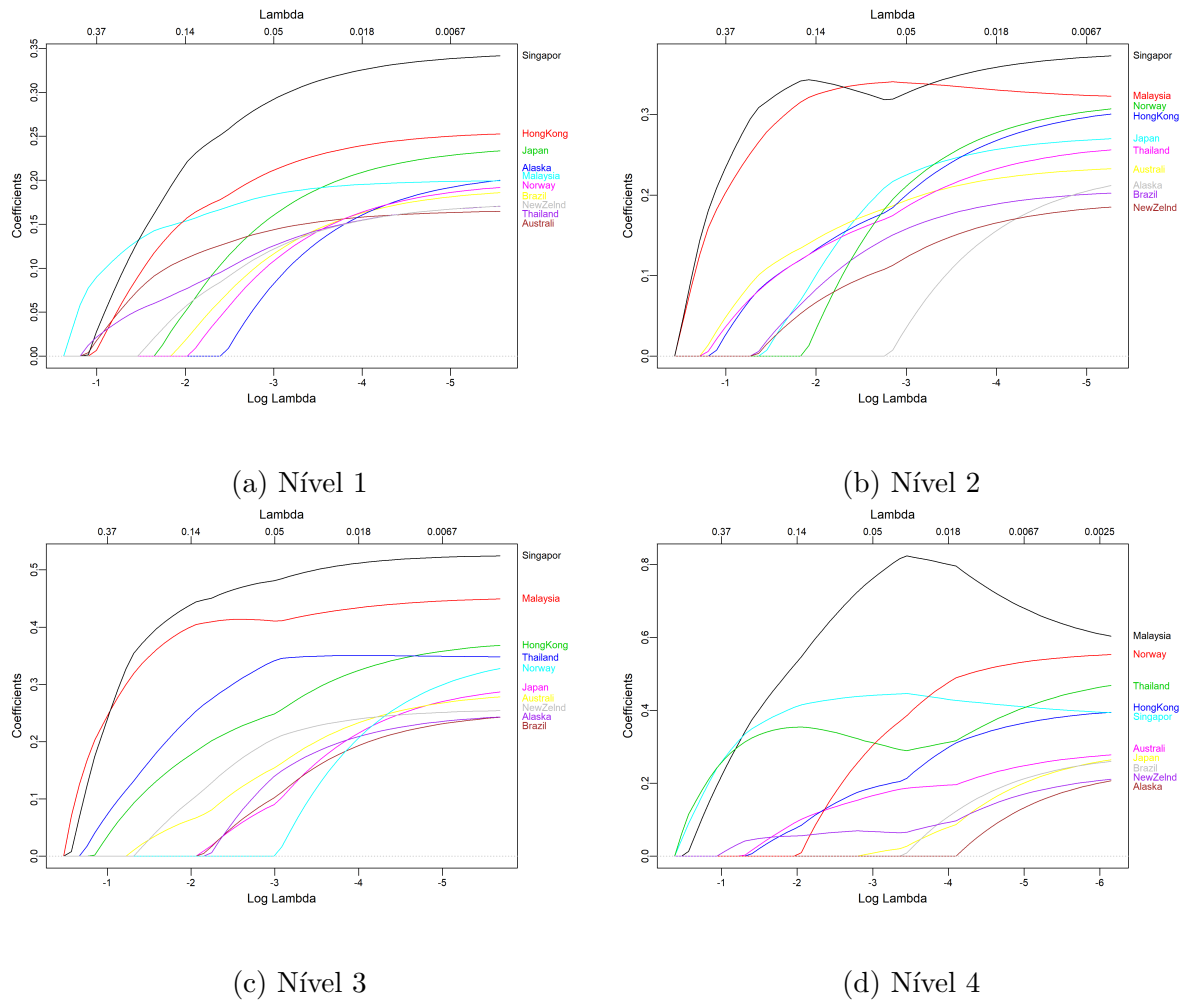


Figura 3: Elastic net para cada nível de detalhe.

Fonte: do autor.

Tabela 2: Grupos obtidos pelo elastic net para cada nível de detalhe.

Níveis	Agrupamentos				
1	{Sin,HK,Tai,Aus}	{Mal}	{NZ,Jap,}	{Bra,Nor}	{Al}
2	{Sin,Mal}	{HK,Tai,Aus}	{NZ,Jap,Bra}	{Nor}	{Al}
3	{Sin,Mal}	{HK,Tai}	{Aus,NZ}	{Jap,Bra,Al}	{Nor}
4	{Sin,Mal,Tai}	{HK,Aus}	{NZ}	{Jap,Bra}	{Nor} {Al}

Fonte: do autor.

Observamos que, a medida que muda-se o nível de detalhe, alguns agrupamentos tendem a se manter enquanto outros rearranjos são evidenciados.

Observamos que a proximidade geográfica entre Singapore, Malaysia, Thailand, Honk Kong e Australia favorece que eles estejam nos mesmos grupos, ainda que alternadamente.

Notamos que Norway e Alaska não se agrupam com nenhum outro país em quase todos os níveis de detalhe, o que também pode ser justificado pelas suas localizações geográficas.

Por fim, Japan, Brazil e New Zealand apresentam-se no mesmo grupo em vários níveis, sendo que Japan e Brazil estiveram juntos em 3 dos 4 níveis de detalhes.

Sáfadi (2014) considerando o mesmo conjunto de dados utilizando Análise de Componentes Independentes (ICA) apresentou alguns resultados semelhantes aos encontrados neste trabalho. Utilizando-se 3, 4 ou 5 componentes independentes, os mesmos grupos Malaysia e Thailand (mesmo grupo no nível 4), e Australia e New Zealand (nível 3) foram formados.

Conclusões

A combinação entre transformada de ondaletas discreta não-decimada combinada com a metodologia Elastic Net para o agrupamento das séries de nível do mar mostrou-se satisfatória, uma vez que foi possível observar diferentes arranjos de grupos nos diferentes níveis de detalhe da decomposição. Além disso, foi possível comparar com resultados observados na literatura.

Agradecimentos

Os autores agradecem o apoio financeiro da Fapemig.

Referências Bibliográficas

- CHURCH; J. A. et al. Estimates of the regional distribution of sea level rise over the 1950-2000 period. *J. Clim.* 17, p.2609-2625, 2004.
- DASGUPTA; S. et al. The impact of sea level rise on developing countries: a comparative analysis. *Clim. Change*, 93, p.379-388, 2009.
- FERREIRA, L. M.; SÁFADI, T.; FERREIRA, J. L. Wavelet-domain elastic net for clustering on genomes strains. *Genetics and Molecular Biology*, Washington, v.41, n.4, p.884-892, 2018.
- MORAIS, T. S. T. de *Estudo temporal do nível médio do mar em diferentes oceanos*. Lavras: UFLA. 2012. 109p.
- PERCIVAL, D. B.; WALDEN, A. T. *Wavelet methods for time series analysis*. Cambridge: Cambridge University Press. 2000. 611p.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- SÁFADI, T. Using independent component for clustering of time series data. *Applied Mathematics and Computation*, 243, p.522-527, 2014.
- SÁFADI, T. Wavelet-domain elastic net for clustering of volatilities. *Internal Journal of Statistics and Economics*, Washington, v.18, n.4, p.73-80, 2017.

SOLOMON; S. et al. Contribution of Working Group I to the Fourth Assessment, Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom, New York, NY, USA, 2007.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B*, Washington, v.67, n.2, p.301-320, 2005.