Clustering and forecasting of dissolved oxygen concentration on a river basin

Marco Costa and A. Manuela Gonçalves

Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, Apartado 473, 3754 – 909 Águeda, Portugal

Departamento de Matemática e Aplicações, Universidade do Minho, Campus de Azurém da Universidade do Minho, 4800-058 Guimarães, Portugal

Abstract

The aim of this contribution is to combine statistical methodologies to classify geographically homogeneous groups of water quality monitoring sites based on similarities in the temporal dynamics of the dissolved oxygen concentration, to obtain accurate forecasts of this quality variable. Our methodology intends to classify the water quality monitoring sites into spatial homogeneous groups, based on the dissolved oxygen concentration, which has been selected and considered relevant to characterize the quality of the water. We apply clustering techniques based on Kullback Information measures which are obtained in the state space modeling process. For each homogeneous group of water quality monitoring sites, we model the dissolved oxygen concentration using linear and state space models which incorporate tendency and seasonality components in different ways. Both approaches are compared by the mean squared error of forecasts.

Keywords: hydrological basin, water quality, clustering, state space model, linear model, Kalman filter.

References

Alpuim T, Barbosa S (1999) The Kalman filter in the estimation of area precipitation. Environmetrics 10:377-394

Bengtsson T, Cavanaugh JE (2008) State-Space Discrimination and Clustering of Atmospheric Time Series Data Based on Kullback Information Measures. Environmetrics 19:103–121

Boi P (2004) A statistical method for forecasting extreme daily temperatures using ECMWF 2-m temperatures and ground station measurements. Meteorol Appl 11:245-251

Brown P, Diggle P, Lord M, Young P (2001) Space-time calibration of radar rainfall data. Appl Statist 50(2):221-241

Carl G, Kühn I (2008) Analysing spatial ecological data using linear regression and wavelet analysis. Stochastic Environmental Research and Research and Risk Assessment, 22(3), 315-324.

Costa M, Alpuim T (2010) Parameter Estimation of State Space Models for Univaritate Observations. J Stat Plan Infer. doi: 10.1016/j.jspi.2010.01.036

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc Ser B 39:1-38

Everitt BS, Landau S, Leese M (2001) Cluster Analysis (4th Ed). Arnold, London

Fovell R, Fovell M (1993) Climate zones of the conterminous United States defined using cluster analysis. J Climate 6:2103-2135

Galanis G, Anadranistakis M (2002) A one-dimensional Kalman filter for the correction of near surface temperature forecast. Meteorol Appl 9:437-441

Gong X, Richman M (1995) On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. J Climate 8:897-931

Harvey AC (1996) Forecasting structural time series models and the Kalman filter. Cambridge University Press: Cambridge

Libonati R, Trigo I, DaCamara C (2008) Correction of 2 m-temperature forecasts using Kalman Filtering technique. Atmos Res 87:183-197

Kullback S. (1968) Information Theory and Statistics. Dover

Leybourne SJ (2006) Estimation and testing of time-varying coefficient regression models in the presence of linear restrictions. J Forecast 12(1):49-62

Mouriño H., Barão, M.I. (2009). A comparison between the Linear Regression Model with autocorrelated errors and the Partial Adjustment Model. Stoch Env Res Risk A. doi: 10.1007/s00477-009-0340-0.

Oliveira RES, Lima MMCL, Vieira JMP (2005) An Indicator System for Surface Water Quality in River Basins. Inter-Celti Colloquium on Hydrology and Management of Water Resources 4 Guimarães

Pagan A (1980) Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients. J Econometrics 13: 341-363

Paschalidou AK, Kassomenos PA, Bartzokas A (2009) A comparative study on various statistical techniques predicting ozone concentrations: implications to environmental management. Environ Monit Assess. 148: 277-289

PGIRH/N (1988) Metodologias para a Avaliação de Políticas de Recursos Hídricos - Plano de Gestão da Bacia Hidrográfica do Rio Ave (in Portuguese). Ministério das Obras Públicas,

Transportes e Comunicações, Laboratório Nacional de Engenharia Civil, Ministério do Plano e Administração do Território, Comissão de Coordenação da Região Norte 8:66, Lisboa

PGIRH/N, NATO PO-RIVERS (1994) Caracterização e Directrizes de Planeamento dos Recursos Hídricos do Norte – A Bacia Hidrográfica do Rio Ave (in Portuguese). Ministério do Ambiente e dos Recursos Naturais, Direcção Regional do Ambiente e Recursos Naturais, Instituto da Água. Porto 1-5;1-13

Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environ Modell Softw 22:464-475

Shumway R, Stoffer D (1982) An approach to time series smoothing and forecasting using the EM algorithm. J Time Ser 3:253-264

Shumway R, Stoffer D (2006) Time Series Analysis and Its Applications (2 ed). Springer-Verlag

Stone RC (1989) Weather types at Brisbane, Queensland: An example of the use of principal components and cluster analysis. Int J Climatol 9:3-32

Vieira JMP (2003) Water Management in National Water Plan Challenges (in Portuguese). Revista Engenharia Civil 16:5-12

Zhu R, El-Shaarawi AH (2009) Model clustering and its application to water quality monitoring. Environmetrics 20:190–205