II Encontro Galaico-Português de Biometria
Santiago de Compostela, 30 de Junho, 1 e 2 de Julho de 2016

STRUCTURAL TIME SERIES MODELING: AN APPLICATION TO ENVIRONMENTAL VARIABLES

Marco Costa[1], Olexandr Baturin[2] and A. Manuela Gonçalves[3]

[1]CIDMA-Centre for Research and Development in Mathematics and Applications, University of Aveiro, Portugal (marco@ua.pt)
[2]DMA-Department of Mathematics and Applications, University of Minho, Portugal (olexandr.baturin@gmail.com)
[3]CMAT-Centre of Mathematics, DMA-Department of Mathematics and Applications, University of Minho, Portugal (mneves@math.uminho.pt)

ABSTRACT

A structural time series model is one which is set up in terms of components which have a direct interpretation. In this paper, the discussion focuses on the dynamic modeling procedure based on the state space approach (associated to the Kalman filter), in the context of surface water quality monitoring, in order to analyze and evaluate the temporal evolution of the environmental variables, and thus identify trends or possible changes in water quality (change point detection). The approach is applied to environmental time series: time series of surface water quality variables in a river basin. The statistical modeling procedure is applied to monthly values of physico-chemical variables measured in a network of 8 water monitoring sites over a 15-year period (1999-2014) in the River Ave hydrological basin located in the Northwest region of Portugal.

Keywords: River Ave, Water quality variables, Structure time series models, State space models, Kalman filter, Change point detection.

1. INTRODUCTION

Structural time series models are formulated directly in terms of unobserved components, such as trends, cycles and seasonals, that have a natural interpretation and represent the salient features of the series under investigation. The key to handling structural time series models is the state space form with the state of the system representing the various unobserved components. State space models provide a very flexible yet fairly simple tool for analyzing dynamic phenomena and evolving systems, and have significantly contributed to extend the classic domains of application of statistical time series analysis to non-stationary, irregular processes, etc. A structural model can therefore not only provide forecasts, but also, through estimates of the components, present a set of stylised facts and this formulation will allow making some useful interpretations [3] and [4]. In this study, it is proposed a dynamic modeling procedure based on the state space approach (associated to the Kalman filter) in time series of water quality variables. The data concerns the River Ave hydrological basin located in the Northwest of Portugal, where monitoring has become

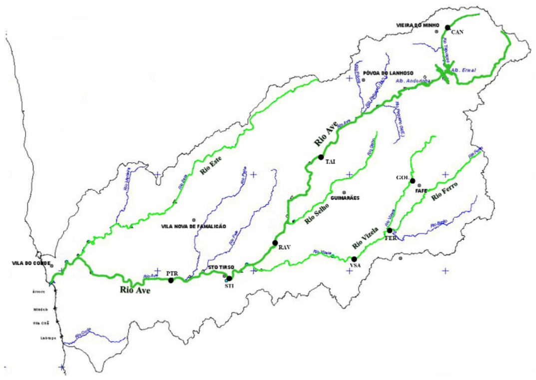

Figure 1: River Ave hydrological basin located in the Northwest of Portugal and its 8 water monitoring sites.

a priority in water quality planning and management because its water has been in a state of obvious environmental degradation for many years. As a result, the watershed is now monitored by eight sampling stations distributed along the River Ave and its main streams: PTR, STI, RAV, VSA, FER, GOL, TAI and CAN (Figure 1). In these water quality monitoring sites are held monthly measurements and analyses for a general assessment of basin's surface water quality (physico-chemical and microbiological variables).

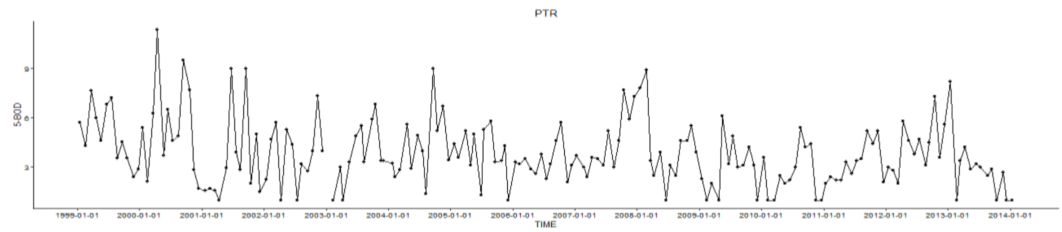

Figure 2: 5-BOD monthly concentration from January 1999 to January 2014 in Ponte Trofa (PTR) monitoring site.

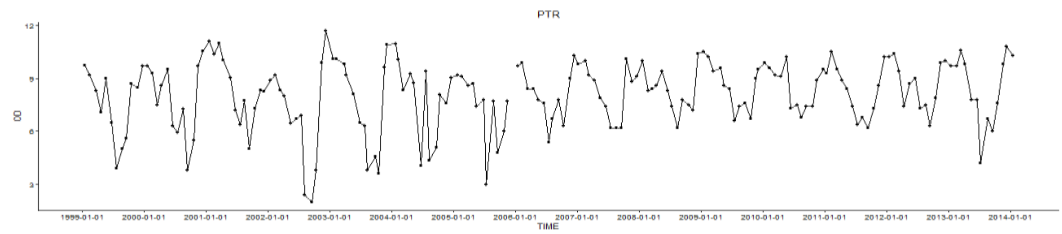

Figure 3: DO monthly concentration from January 1999 to January 2014 in Ponte Trofa (PTR) monitoring site.

For this study we selected the Dissolved Oxygen (OD) concentration and the 5-day Biological Oxygen Demand (5-BOD) due to their importance in the evaluation of this river's water quality, both measured in mg/l. Dissolved oxygen (OD) refers to the level of free, non-compound oxygen present in water or other liquids. It is an important variable in assessing water quality because of its influence on the organisms living within a body water. Dissolved oxygen (OD) can range from less than $1mg/l$ to more than $20mg/l$. Biochemical Oxygen Demand (5-BOD) refers to the amount of oxygen that would be consumed if all the organics in one liter of water were oxidized by bacteria and protozoa. Microorganisms such as bacterias are responsible for decomposing organic waste. When organic matter such as dead plants, sewage, etc. is present in a water supply, the bacteria will begin the process of breaking down this waste. Biochemical Oxygen Demand (BOD) is a measure of the oxygen used by microorganisms to decompose this waste. The range of possible readings can vary considerably: water from an exceptionally clear water might show a 5-BOD of less than $2mg/l$ of water. In this paper we consider data series from one water monitoring site PTR (Ponte Trofa) on pretence of example of the modeling approach (Figures 2 and 3). For the modeling process we consider data between January 1999 to January 2014 ($n = 181$ observations).

2. MODELING APPROACH

An exploratory analysis was performed and showed that the data exhibit seasonal behavior (by analyzing the monthly averages) and a moderate temporal dependency [2] and [3]. The linearity of the possible trend that was suggested in [2] by a graphical analysis of the 12 times series corresponding to each month was found not suitable for the time series. However, we must explore a stochastic level as p-dimensional state space in the modeling. The parameters of the state space models must be estimated for each environmental series, and they are estimated by Gaussian maximum likelihood estimation in the state space framework. As the series of observations present intrinsic environmental data proprieties, the initial model is very versatile since it can accommodate several statistical properties often presented in environmental data, such as pth order polynomial component (a stochastic local level is considered), seasonality and temporal correlation. It is proposed an application of a structural time series model by taking into account this data structure. So, the monthly time series are modeled by equations

$$y_t = \mu_t + \ _t + \ _t$$

$$\mu_{t+1} = \mu_t + \ _t$$

$$\ _{1,t+1} = - \ _{1,t} - \ _{2,t} - \ _{3,t} \cdots - \ _{11,t} + \ _t$$

$$\ _{2,t+1} = \ _{1,t}, \ \ _{3,t+1} = \ _{2,t}, \cdots, \ \ _{11,t+1} = \ _{10,t}$$

$$\ _t = \ _1 \ _{t-1} + \ _t$$

with

$$\ _t \ \ N(0, \ ^2), \ \ _t \ \ N(0, \ ^2) \text{ and } \ _t \ \ N(0, \ ^2).$$

The first equation represents the structure for the observed data, the observation equation, where y_t is the monthly environmental variable, in this case 5-BOD and OD variables with $t = 1, 2, ..., 181$. The observations are driven by two state variables, one being the level variable and the second being the seasonality variable. The second equation is a state space equation representing a stochastic level as an 1-dimensional state space. The seasonal behavior is represented by $\ _t$, an 11-dimensional state space vector of seasonal deviations. And to deal with the temporal correlation structure it is considered $\ _t$ following a 1st order autoregressive process, AR(1). Since the model has a state space representation, it allows obtaining forecast or other predictions of interest (filtered or smoother predictions). Since the state process is unobserved, both forecast and filtered predictions are obtained through the Kalman filter algorithm.

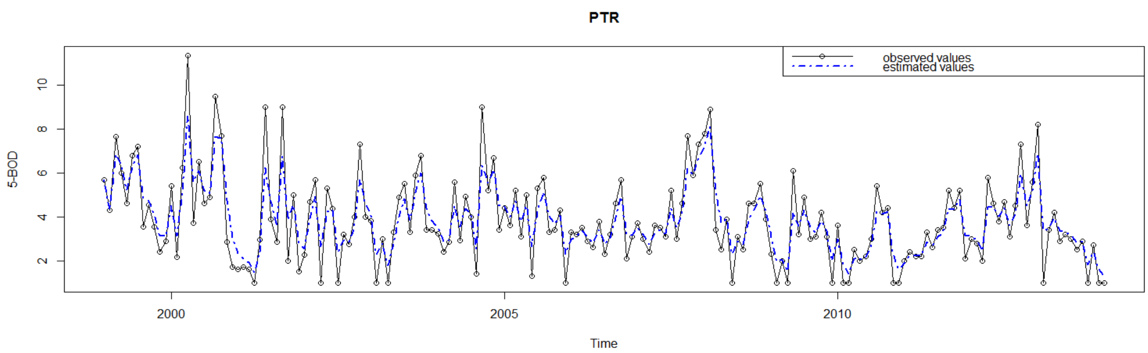

Figure 4: Observed values and filtered state estimates of 5-BOD in PTR monitoring site.

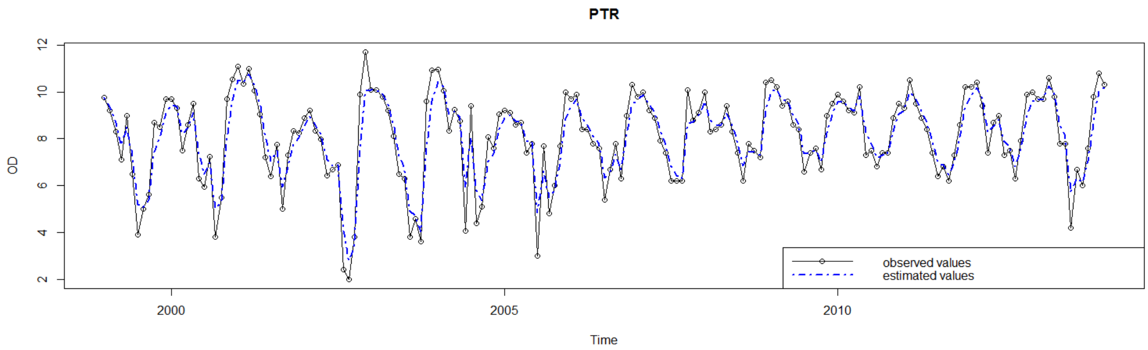

Figure 5: Observed values and filtered state estimates of OD in PTR monitoring site.

The referred methodology was applied in R environment with dlm package (http://www.r-project.org/) [5] and [6]. A first analysis of adjustment models is done by comparing data and predictions. Figures 4 and 5 show observed data and the filtered predictions in Ponte Trofa (PTR) for 5-BOD and OD.

3. CONCLUSIONS

The analysis performed in this study shows that the structural time series models (the state space models associated to the Kalman filter) are suitable to model 5-BOD and OD concentration series at the PTR water monitoring site and they allow to obtain pertinent findings concerning water surface quality interpretation and of change point of view [1] and [4], thus highlighting the potential value of this type of analysis. The next step is to analyze the filtered and smoothed predictions (forecasts) of states given by the Kalman filter, which allows an interesting analysis of the structural components, and further research is in progress to improve the modeling process and the results obtained.

ACKNOWLEDGEMENTS

Marco Costa was supported by Portuguese funds through the CIDMA-Centre for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology "FCT-Fundação para a Ciência e a Tecnologia", within project UID/MAT/04106/2013. A. Manuela Gonçalves was supported by the Research Centre of Mathematics of the University of Minho with the Portuguese Funds from the "FCT-Fundação para a Ciência e a Tecnologia", through the Project PEstOE/MAT/UI0013/2014.

References

[1] Costa M., Gonçalves A.M. (2011). Clustering and forecasting of dissolved oxygen concentration on a river basin. *Stochastic Environmental Research and Risk assessment* 25(2):151–163.

[2] Gonçalves A.M., Alpuim T. (2011). Water Quality Monitoring using cluster analysis and linear models. *Environmetrics* 22(8), 933–945.

[3] Gonçalves A.M., Costa M. (2013). Predicting seasonal and hydro-meteorological impact in environmental variables modelling via Kalman filtering. *Stochastic Environmental Research and Risk assessment* 127(5):1021–1038.

[4] Gonçalves A.M., Costa M. (2011). Application of Change-Point Detection to a Structural Component of Water Quality Variables. In Simos, T.E., Psihoyios, G., Tsitouras, Ch. (eds.): *AIP Conference Proceedings* Vol 1389, 1565–1568, AIP American Institute of Physics Publishing.

[5] Petris, G., Petrone, S., Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer, New York.

[6] R Team, D. (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.