Poster Presentation

Improvement of surface water quality variables modelling that incorporates a hydro-meteorological factor: a state-space approach

A. Manuela Gonçalves[1], Marco Costa[2]

[1] Department of Mathematics and Applications, University of Minho
CMAT-Center of Mathematics, Portugal
mneves@math.uminho.pt
[2] Higher School of Technology and Management of Águeda-University of Aveiro
CMAF-UL, Portugal
marco@ua.pt

Abstract: In this work it is constructed a hydro-meteorological factor to improve the adjustment of statistical time series models, such as state space models, of water quality variables by observing hydrological series (recorded in time and space) in a River basin. The hydro-meteorological factor is incorporated as a covariate in multivariate state space models fitted to homogeneous groups of monitoring sites. Additionally, in the modelling process it is considered a latent variable that allows incorporating a structural component, such as seasonality, in a dynamic way.

Keywords: hydrological basin; water quality, state-space modelling; Kalman filter; hydro-meteorological factor.

1 Introdution

Water quality monitoring is an important tool in the management and assessment of surface water quality. This study focuses on a rather extended data set relative to the River Ave basin (Portugal) and consists mainly of monthly measurements of biochemical variables in a network of monitoring water quality stations.A hydro-meteorological factor is constructed for each monitoring station based on monthly estimates of precipitation obtained by means of a rain gauge network. Through stochastic interpolation (Kriging) it is estimated the mean area rainfall during each month in the area of influence of each water quality monitoring site. These estimates are based on rain gauges located in the respective area of influence. In a recent work, Costa and Gonçalves (2010) show that a set of water quality monitoring sites can be modelled applying cluster techniques that minimize the number of models.

2 Data Set Description

The Northern Regional Directory for the Environment and Natural Resources (DRARN) and the National Institute of Water (INAG) has been collecting various water quality variables (monthly physical-chemical and microbiological analyses) from 16 quality monitoring sites. The data set of the 16 water quality monitoring sites, comprising 11 water quality variables, have been monthly measured between 1988 and 2006. At this time, this work focuses on Dissolved Oxygen (DO) (mg/l) in water because it is one of the most important variables in the evaluation on river water quality. For instance, it is shown the data and the results of one cluster with five water monitoring sites identified in Costa and Gonçalves (2010) as the less polluted cluster.

3 Methods

As starting point, it is constructed a hydro-meteorological factor used as covariate in the modelling process. This covariate will integrate a hydro-meteorological component that is recognized as crucial in any water quality modelling process. This factor is constructed through stochastic interpolation (Kriging) based on an udometric network (Figure 1) with 19 meteorological stations. The model of spatial continuity, which is inferred from monthly precipitation estimates, assumes hypothesis of homogeneity of the process: the process is stationary of 2nd, i.e., intrinsically stationary and isotropic. Under this hypothesis, two observations in the same location but in different times are independent and the spatial variability pattern remains the same (Kyriakidis and Journel, 1999). The empirical semivariogram is given by

$$\hat{\gamma}_z(h \mid l) = \frac{1}{2T|N(h|l)|} \sum_{t=1}^{T} \sum_{(i,j)\in N(h|l)} [(Z_t(s_i) - Z_t(s_j)]^2$$

with $N(h|l) = \{(i,j) : \|s_i - s_j\| - \|h\| \leq l; 1 \leq i \leq j \leq n\}$ and $|N(h|l)| = \#N(h|l)$.

The river basin is discretized in 368 points with $2Km$ x $2Km$ (Figure 1) and at each point s_0 the estimate of the monthly mean area precipitation is given by the Kriging estimator, i.e., by a linear combination of the 19 known points s_j, $j = 1, ..., 19$ and $Z_t(s_0) = \sum_{j=1}^{19} \lambda_j Z_t(s_j)$.

3.1 Hydro-meteorological factor

It is constructed one covariate for each water monitoring site based on the estimate of the monthly mean precipitation of its influence region. In this

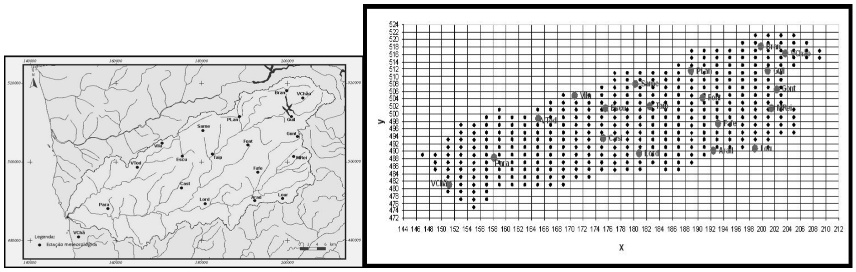

FIGURE 1. Spatial distribution of 19 meteorological monitoring sites in the River Ave basin and discretization of River Ave basin in 368 points.

context, the influence regions of each water monitoring site were defined by technicians of the INAG and they are supported on the regions topography and the lands drainage dynamics.

Firstly, for each water monitoring site, it was computed the monthly mean area precipitation in its influence region based on the average of point prediction. Naturally, a large influence region tends to have a greater precipitation amount. Indeed, it is clear that the precipitation amount influences oxygen concentration in water. However, if the goal of this work is to found a prediction model to DO in a month t, the covariate should not incorporate the precipitation amount of the current month, but only the past information.

Let $P_t^{(i)}$ be the estimate of the precipitation amount in the influence area of a water monitoring site i at month t. We considered a covariate $H_t^{(i)}$computed as a weighted average of precipitation amount at months $t-1$ and $t-2$.

3.2 State space model

For each cluster i with homogenous water monitoring site it is fitted a state space model to Dissolved Oxygen concentration incorporating two structural components: the hydro-meteorological factor and a seasonality. In order to simplify, it is considered monthly seasonality assuming 12 known coefficients (for each month it is taken the month mean; Costa and Gonçalves, 2010):

$$\begin{pmatrix} Y_{1,t}^{(i)} \\ Y_{2,t}^{(i)} \\ \vdots \\ Y_{m,t}^{(i)} \end{pmatrix} = \begin{pmatrix} S_t & H_{1,t}^{(i)} \\ S_t & H_{2,t}^{(i)} \\ \vdots & \vdots \\ S_t & H_{m,t}^{(i)} \end{pmatrix} \begin{pmatrix} X_{1,t}^{(i)} \\ X_{2,t}^{(i)} \end{pmatrix} + \begin{pmatrix} \mu_{1,t}^{(i)} \\ \mu_{2,t}^{(i)} \\ \vdots \\ \mu_{m,t}^{(i)} \end{pmatrix},$$

where

$$\begin{pmatrix} X_{1,t}^{(i)} \\ X_{2,t}^{(i)} \end{pmatrix} = \begin{pmatrix} 1 \\ \mu_{X_2}^{(i)} \end{pmatrix} + \begin{pmatrix} \phi_{11}^{(i)} & \phi_{12}^{(i)} \\ \phi_{21}^{(i)} & \phi_{22}^{(i)} \end{pmatrix} \left[\begin{pmatrix} X_{1,t-1}^{(i)} \\ X_{2,t-1}^{(i)} \end{pmatrix} - \begin{pmatrix} 1 \\ \mu_{X_2}^{(i)} \end{pmatrix} \right] + $$
$$+ \begin{pmatrix} V_{1,t}^{(i)} \\ V_{2,t}^{(i)} \end{pmatrix}.$$

Since normal distribution is not always the best distribution to fit meteorological variables in this work, we adopted consistent distribution-free estimators developed from the original work by Costa and Alpuim (2010).

TABLE 1. Parameters estimates.

$\hat{\mu}_X$	$\hat{\phi}$		$\hat{\Sigma}_V$		$\hat{\Sigma}_\mu$					*Sites*
1	0.277	−1.045	0.016	−0.005	0.597	0.000	0.000	0.000	0.000	*CANT*
−0.0003	0.038	0.738	−0.005	0.003	0.000	0.265	0.000	0.000	0.000	*GOL*
					0.000	0.000	0.417	0.000	0.000	*FER*
					0.000	0.000	0.000	0.383	0.000	*VSA*
					0.000	0.000	0.000	0.000	0.737	*TAI*

The state space model with these parameters estimates associated to the Kalman filter produces monthly one-step predictions for Dissolved Oxygen concentration at each water monitoring site (Table 1). Figure 2 shows observed data and predictions in Vizela Santo Adrião (VSA) and Golães (GOL) monitoring sites.

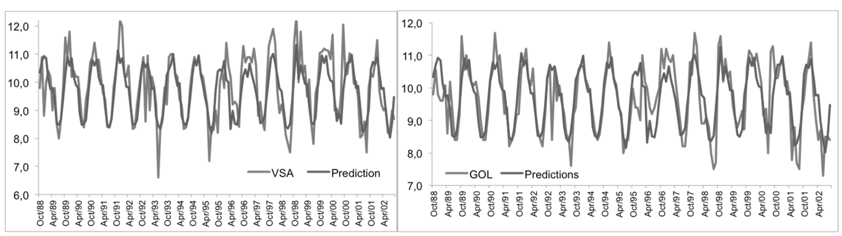

FIGURE 2. Observed and one-step predictions of Dissolved Oxygen concentration in Vizela Santo Adrião (VSA) and in Golães (GOL).

4 Conclusions

It is possible to conclude that the hydro-meteorological factor is an important component adding information beyond the usual seasonality. Moreover, the adoption of the consistent distribution-free estimators for the state space models requires a future comparison with gaussian likelihood estimation, assessing its relative efficiency, and possibly comparing its forecasts mean square error. However, distribution-free estimators are an easy solution without computed problems, nor iterative procedures and neither requires initial values. The next step is to analyse the filtered estimates of states $X_{t|t}^{(i)}$ given by the Kalman filter, which allows an interesting analysis of these latent variables as calibrate factors of the two structural components.

Acknowledgments: A. Manuela Gonçalves acknowledges the financial support provided by the Research Centre of Mathematics of the University of Minho through the FCT Pluriannual Funding Program.

References

Bengtsson, T., Cavanaugh, J. (2008). State-space discrimination and clustering of atmospheric time series data based on Kullback information measures. *Environmetrics*, **10**, 377-394.

Costa, M., Alpuim, T. (2010). Parameter estimation of state space models for univariate observations. *J Stat Plan and Inference*, 140**(7)**, 1889-1902.

Costa M., Gonçalves, A. M. (2010). Clustering and forecasting of dissolved oxygen concentration on a river basin. *Stoch Environ Res Risk Assess*, DOI: 10.1007/s00477-010-0429-5.

Kyriakidis, P.C., Journel, A.G. (1999). Geostatistical space-time models: a review. *Mathematical Geology*, 31**(6)**, 651-684.