# Statistical Modeling in the Analysis of River Water Quality Monitoring Sites

Carla Sofia Silva[1], A. Manuela Gonçalves [2] and Susana Faria[3]

[1] Department of Mathematics, University of Minho, Guimarães, Portugal
[2] Department of Mathematics, CMAT - Centre of Mathematics, University of Minho, Guimarães, Portugal
[3] Department of Mathematics, CBMA - Centre of Molecular and Environmental Biology, University of Minho, Guimarães, Portugal

E-mail for correspondence: mneves@math.uminho.pt

**Abstract:** Surface water quality monitoring has as its main objective the characterization of water resources as well as the monitoring of its space-time evolution in order to achieve an appropriate administration. Dissolved oxygen (DO) is an important indicator for the estimation of the water quality index, and the lack of this quality variable may cause significant environmental issues. The aim of this contribution is to propose a geostatistical model that can used to characterize and assess water quality behaviour/evolution based on a rather extended data relative to the River Douro Basin (in Portugal) that consists mainly of monthly measurements of dissolved oxygen concentration in a network of water quality along this basin. The monthly data set reports to the period between March 2002 and February 2013.

**Keywords:** Geostatistics; River Basin; Spatio-Temporal Models; Water Quality.

## 1 Introduction

The degradation of water resources due to pollution from anthropogenic activities (van Dijk et al., 1994) is undeniable. There is a lack of normalized methods, management plans and tools that forecast critical events in order to preserve water quality. In a holistic way, a sequential modeling process was developed to forecast the status of the river basin by stochastic dynamic method (Cabecinha et al., 2009, Silva-Santos et al., 2008) and, in particular to predict DO as a measure of water quality.

---

## 2 Data Set

The National Information System on Water Resources (SNIRH) is responsible for gathering several environmental variables, namely water quality variables. The data has several variables, yet only the dissolved oxygen (DO) variable is chosen owing to its continuity characteristics, its importance (since most aquatic fauna and flora need oxygen to survive) and its localization in key points. This study focus on monthly measurements of DO in 36 quality monitoring sites located along the main course of river Douro in Portugal, from March 2002 to February 2013 (Figure 1).

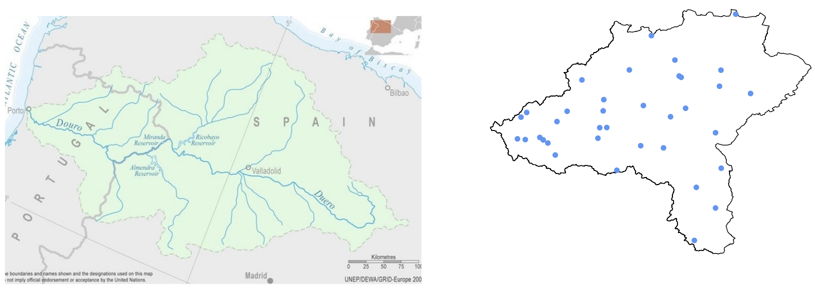


FIGURE 1. The Douro river hydrological basin (left) and spatial distribution of the water quality monitoring sites (right).

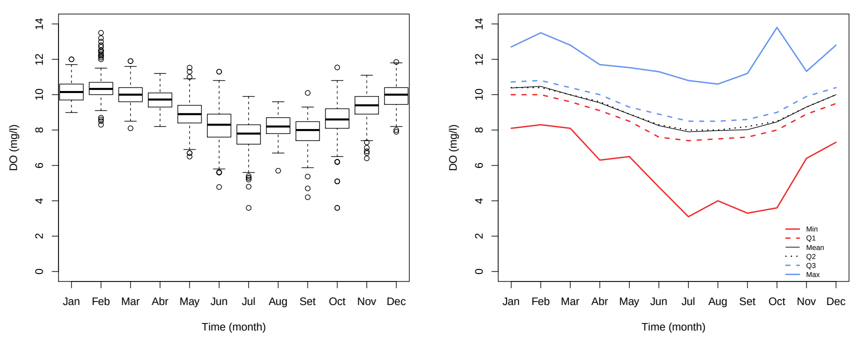


FIGURE 2. Boxplot (left) and descriptive measures (right) of dissolved oxygen (DO) by months during the observed period.

The Douro river source is located in the Spanish mountains of Urbión and it flows into the Atlantic Ocean, alongside Oporto city, with 927 km (330 km in Portugal) of length and 97667 km$^2$ of hydrological basin area. The Douro river has a basin slope average of 9.40%, annual flow average of 903

$m^3$/s, annual temperature average of 13.50°C and annual rainfall average of 908 mm. A preliminary analysis of data reveals a periodicity, seasonal and spatial patterns. It is clear that DO behaviour depends on season (there is a higher DO concentration in winter/spring than in fall/summer, Figure 2) and on the localization (DO concentration in the countryside is lower than at the seaside).

## 3 Methodology

The initial concept is a geostatistics analysis of space by applying the basic concepts of Kriging and its techniques (Cressie, 1993). DO concentration recorded over time shows that there is a seasonal pattern. This fact leads to an approach that consists of the following: the observed DO concentration along the years were separated according to the twelve months. For every month, the spatial continuity of DO concentration was examined to evaluate the temporal component. This method is applied to a data set of dissolved oxygen measurements collected in 36 different and independent monitoring sites locates along the River Douro. No-transformation of DO concentration data was also considered. The models of spatial continuity were inferred from the monthly DO concentration. The empirical semivariograms were obtained by using the method of moments, modified for a random space-time process. For each month, the empirical semivariograms were calculated as well as the number of data pairs that are needed to estimate. For instance, only the results of applying the semivariogram corresponding to the month of January (of the 12 months of the year) are presented, Figure 3. We estimate the values and the least squares adjustments to several stationary models have been performed.

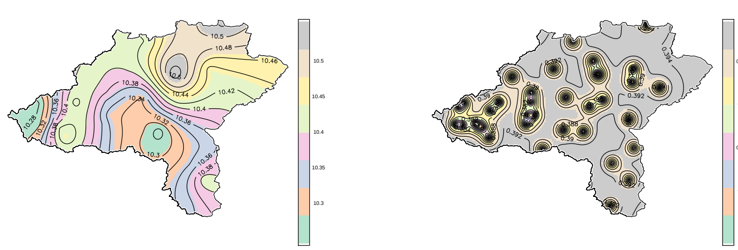


FIGURE 3. Plots of the experimental estimator and the fitted model (left) and standards errors (right) for the spatial semivariogram.

This methodology could be further developed to better fulfill other applications requirements, such as other water quality variables. Spatio-temporal models have a potential to overtake the usual linear regression model in terms of its ability to integrate the temporal dynamic intrinsic to the water quality monitoring process. The results of applying to a real database should be created and used to compare with previous studies results.

**Acknowledgments:** FEDER/COMPETE/- NORTE2020/ POCI/ FCT funds through grants PTDC-EEI-AUT-2933-2014116858-TOCCATA and To CHAIR - POCI-01-0145-FEDER-028247 Financial support from the Portuguese Foundation for Science and Technology (FCT) within the framework of the Strategic Financing UIDIFIS/04650/2013 is also acknowleged. The research of the A. Manuela Gonçalves author was partially financed by Portuguese Funds through FCT (Fundação para a Ciência e a Tecnologia) within Project UID/MAT/00013/2013. This work was partially supported by the R&D Project "PREFERENTIAL"-PTDC/MAT-STA/28243/2017, sponsored by the Foundation for Science and Technology, I.P. (FCT, IP).

## References

Cabecinha, E., Cortes, R., Pardal, M. Â. and Cabral, J. A. (2009). A Stochastic Dynamic Methodology (StDM) for reservoir's water quality management: Validation of a multi-scale approach in a south european basin (Douro, Portugal). *Ecological Indicators*, **9(2)**, 329 – 345.

Costa, M. and 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.

Cressie, N.A.C. (1993). Statistics for Spatial Data. *2nd edition. Wiley, New York.*

Gonçalves, A. M. and Alpuim, T. (2011). Water quality monitoring using cluster analysis and linear models. *Environmetrics*, **22(8)**, 933 – 945.

Silva-Santos, P., Pardal, M. Â., Lopes, R. J., Múrias, T. and Cabral, J. A. (2008). Testing the Stochastic Dynamic Methodology (StDM) as a management tool in a shallow temperate estuary of south Europe (Mondego, Portugal). *Ecological Modelling*, **210(4)**, 377 – 402.

van Dijk, G. M., van Liere, L., Admiraal, W., Bannink, B. A. and Cappon, J. J. (1994). Present state of the water quality of european rivers and implications for management. *Science of the Total Environment*, **145(1-2)**, 187 – 195.