Science of the Total Environment 605–606 (2017) 276–290

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

# Spatial prediction of water quality variables along a main river channel, in presence of pollution hotspots

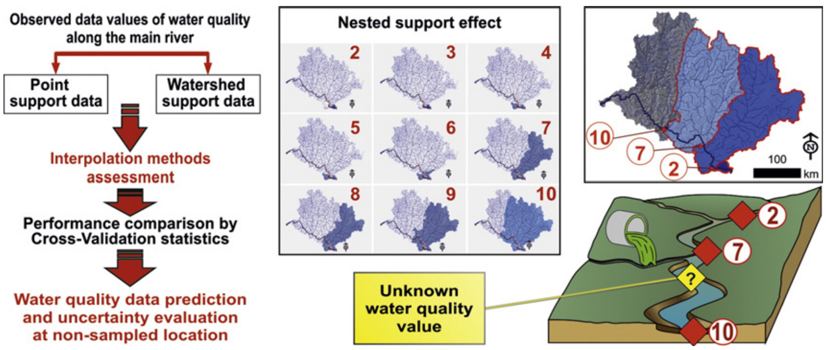
L.D. Rizo-Decelis [a,*], E. Pardo-Igúzquiza [b], B. Andreo [a]

[a] *Universidad de Málaga, Centre of Hydrogeology (CEHIUMA), Faculty of Sciences, Department of Geology, Campus de Teatinos s/n, 29071 Malaga, Spain*
[b] *Instituto Geológico y Minero de España (IGME), Department of Planning and Geosciences Research, Ríos Rosas 23, 28003 Madrid, Spain*

## HIGHLIGHTS

- An approach for estimation of water quality variables along a river is proposed.
- Each river sub-basin area is relevant to predict water-quality variables downstream.
- Different interpolation methods of water-quality variables are assessed along a river.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

*Article history:*
Received 10 March 2017
Received in revised form 14 June 2017
Accepted 18 June 2017
Available online 29 June 2017

Editor: D. Barcelo

*Keywords:*
Geostatistics
River pollution
Water quality
Interpolation
Spatial analysis
Mexico

## ABSTRACT

In order to treat and evaluate the available data of water quality and fully exploit monitoring results (e.g. characterize regional patterns, optimize monitoring networks, infer conditions at unmonitored locations, etc.), it is crucial to develop improved and efficient methodologies. Accordingly, estimation of water quality along fluvial ecosystems is a frequent task in environment studies. In this work, a particular case of this problem is examined, namely, the estimation of water quality along a main stem of a large basin (where most anthropic activity takes place), from observational data measured along this river channel. We adapted topological kriging to this case, where each watershed contains all the watersheds of the upstream observed data ("nested support effect"). Data analysis was additionally extended by taking into account the upstream distance to the closest contamination hotspot as an external drift. We propose choosing the best estimation method by cross-validation. The methodological approach in spatial variability modeling may be used for optimizing the water quality monitoring of a given watercourse. The methodology presented is applied to 28 water quality variables measured along the Santiago River in Western Mexico.

## 1. Introduction

Although fresh water resources affect every human activity, aquatic ecosystems are among the most endangered on Earth (Nel et al., 2009). Deterioration of rivers and streams due to human activities is a critical issue (Namour et al., 2015), yet water quality monitoring procedures, for river conservation and management, are still limited (Isaak et al., 2014).

---

* Corresponding author.
*E-mail address:* rizo@uma.es (L.D. Rizo-Decelis).
*URL's:* (L.D. Rizo-Decelis), http://www.igme.es/ (E. Pardo-Igúzquiza), http://www.cehiuma.uma.es/en/ (B. Andreo).

http://dx.doi.org/10.1016/j.scitotenv.2017.06.145

For most large river basins, insufficient data are collected in situ (e.g. hydrochemical, microbiological, and other physico-chemical information) to report back about surface water quality and its contamination. Publicly available information about river water contamination is often limited to the legislated pollutants only, while few important variables might be measured and/or the number of sampling stations is considerably restricted (Ani et al., 2011). Thus, estimation of water quality along a river network is challenging — particularly due to the scarcity of sampling locations — while also problematic for water resources management. More relevant information could aid in making better decisions about stream water assessment and management, helping to identify contamination sources, and providing insight for the location and redesign of sampling campaigns (Yang and Jin, 2010). It is essential to develop improved and efficient methodologies to treat and evaluate the available data (Isaak et al., 2014; Álvarez-Cabria et al., 2016).

In recent times, Geostatistics has become consolidated as a useful approach for predicting the spatial and temporal variability of different water quality parameters (Goovaerts, 1997; Garreta et al., 2010; Morio et al., 2010), and for improving monitoring techniques (Meyer et al., 2015). It is crucial to characterize regional water quality patterns, and optimize monitoring networks, to fully exploit the available monitoring data results, and to infer water quality conditions at unmonitored locations. In the unlikely case of pristine zones (i.e. sectors with no anthropogenic alteration related to water pollution) in a river basin, Topological Kriging (TK) also known as Top-kriging, first proposed by Skøien et al. (2006), has proven to be an optimal methodology for estimating streamflow-related variables along streams (Laaha et al., 2012, 2013, 2014). Mean annual discharge and concentration of pollutants are some key variables. Most often, pristine conditions have been largely lost, as human activities contaminate the river networks. The streamflow-related behavior of concentration is therefore not fully met in TK, and must be modified accordingly.

Spatial statistics on stream networks represent an active research area in environmental statistics (Ver Hoef et al., 2006; Isaak et al., 2014). Its purpose is to improve predictions and make estimations closer to real data measured in situ. Classical geostatistical solutions for interpolation by kriging (Goovaerts, 1999), and for network optimization (Pardo-Igúzquiza, 1998), are based on Euclidean distance between the observed data and the unmonitored locations. Spatial statistics on stream networks may consider one or several of the seven following aspects of stream topology:

i. Using stream distances instead of Euclidean distances (Ver Hoef et al., 2006)
ii. Consider every observation location not as a measurement with point support, but as areal support, which is equal to the watershed draining to that point (Skøien et al., 2014)
iii. Watersheds having a hierarchical and nested structure (Isaak et al., 2014)
iv. New models of covariance valid for stream networks (Laaha et al., 2012; Müller and Thompson, 2015)
v. New connectivity definitions (Skøien et al., 2006)
vi. Directionality in the definition of connectivity or distances (Brammer, 2014)
vii. Consider the pollution hotspots (Tsuzuki, 2015).

When water quality estimation adopts the concept of nested support — i.e. a basin that contains a smaller basin of the same type inside, which has, in turn, another basin inside of it, and so on — in watershed support areas, it may allow for more accurate prediction of pollutant concentration in rivers. On one hand, this model considers both the draining surface and its influence on dilution processes, which are deeply involved in the natural attenuation capacity of rivers (Chang, 2008; Tsuzuki, 2015). On the other hand, the location of water pollution-hotspots (i.e. where wastewater discharges from specific sources occur, and may expose the river to elevated and localized pollutants concentration) are also taken into account. Possibly, the most obvious stream variable is runoff (Müller and Thompson, 2015), however, since there is a correlation between flow-rate and the dilution capacity of streams, there are many other variables related to water-flow, such as the measurement of water physicochemical variables, concentration of chemical elements, and microbiological indicators.

We hypothesize that Top-kriging (TK), Top-kriging with external drift (TKED), ordinary kriging (OK), regression kriging (RK), or any combination of these, with respect to distances to the pollution hotspots, will cover a wide range of underlying conditions to assess the estimations precision. The accuracy of the method results will depend on the prediction variability of each water quality variables observed (in the case study, 28 are determined). It can be identified by cross-validation, which is the standard procedure in Geostatistics (Stone, 1974; Bradley, 1983; Chiles and Delfiner, 2012).

The main purpose of this paper is to offer a more accurate methodological approach than the most employed procedures to estimate the water quality, along the main channel of the Santiago River in Mexico, using: (1) the available physicochemical data, (2) the recognized pollution hotspot locations, and (3) the watershed delineation from digital terrain models. Another goal is to display specific results, whose analysis can help optimize the current monitoring procedure of the river water quality.

## 2. Methodology

### 2.1. Study area

The study area is located in the Central-Western region of Mexico. It covers the first 281.5 km stretch of the Santiago River, from the river-source to its confluence with the Bolaños River on the boundary of Jalisco and Nayarit states, which represents a catchment area of 52,615.5 km$^2$ (Fig. 1).

The climate in the study area is mainly warm subtropical, with a mean temperature from 18 to 22 °C, characterized by heavy rains in summer (June–September) and relatively warm winters (December – March). Precipitation increases in downstream direction, from 500 to 800 mm/y in headwaters to 800–1400 mm/y in the lower part of the basin, close to La Yesca dam (Fig. 1; SMN, 2015).

Santiago River is about 562 km long. Since 1970, it stems from the NE part of Lake Chapala (with a surface area of ~1100 km$^2$) by artificial pumping due to lower water-levels of the lake in the eastern side (Herdendorf, 1982; De Anda et al., 1998, 2000). It drains ~250 m$^3$/s into the Pacific Ocean, from an altitude of 3140 m to sea level. The wide variety of geological features gives rise to prominent changes in topography of the canyons in the watershed, soil types, and landscape diversity (Moore et al., 1994). The prevailing lithological materials are Cenozoic volcanic rocks (Tertiary), and a small percentage of alluvial material from the Quaternary (Ferrari et al., 1999).

The main land use in the study area is grassland and scrub (40%), forests (30%), rainfed agriculture and livestock (28%), and urban-industrial (2%), according to available mapping charts (INEGI, 2015). The region faces a water crisis coupled with excessive population growth (over ten times in the last six decades). Urbanization and the installation of industrial facilities in the absence of planning strategies and proper wastewater treatment have resulted in deterioration of the Santiago River water quality (IMDEC, 2007). Mexican water management has been focused on the construction of major infrastructure for distribution and sanitation (CONAGUA, 2015), with a lack of implementation of adequate pollution-control strategies (Rojas-Ortuste, 2014). Consequently, most of the surface waters in the basin of the Santiago River are contaminated.

Lake Chapala represents the primary source of drinking water for Guadalajara city (Fig. 1), home to over 4.5 million people. Yet paradoxically, the lake receives a high amount of wastewater discharge from the densely populated area of Toluca Valley, west of Mexico City, through