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

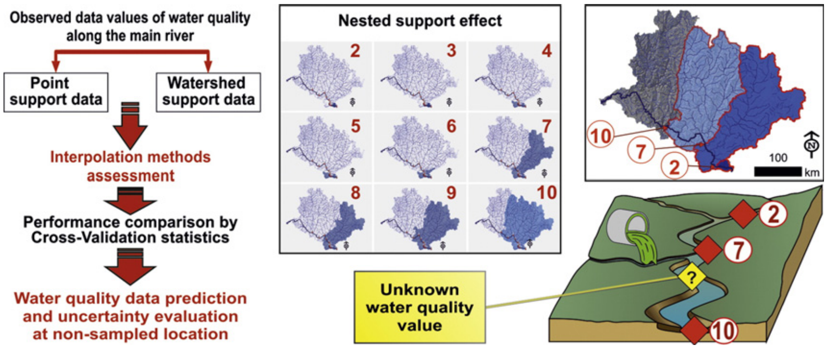
L.D. Rizo-Decelis [a,*], E. Pardo-Igúzquiza [b], B. Andreo [a]

[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


Observed data values of water quality along the main river
Point support data
Watershed support data
Interpolation methods assessment
Performance comparison by Cross-Validation statistics
Water quality data prediction and uncertainty evaluation at non-sampled location
Nested support effect
2
3
4
5
6
7
8
9
10
10
7
2
100
km
N
2
7
?
10
Unknown water quality value

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
0048-9697/© 2017 Elsevier B.V. All rights reserved.

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

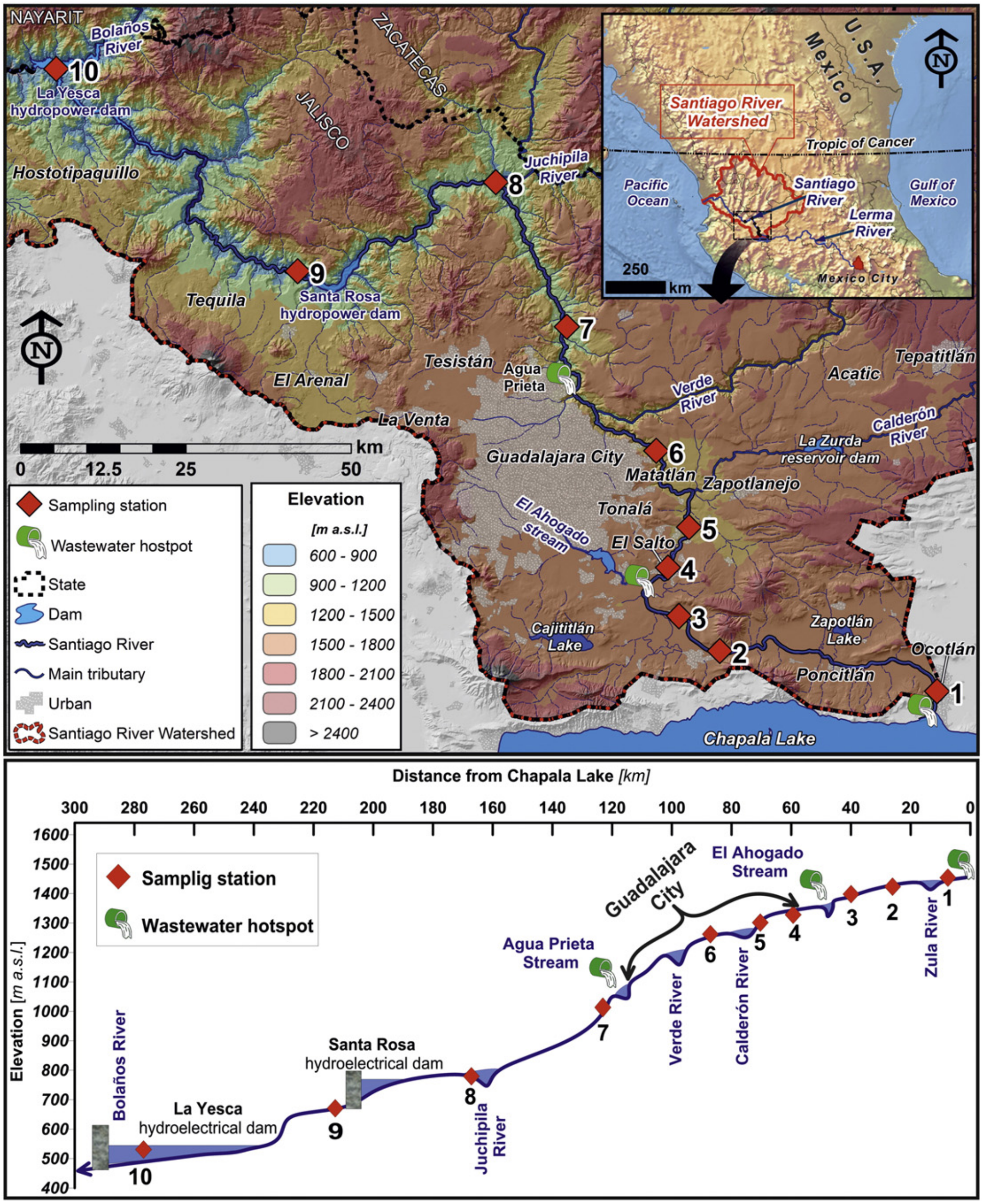


Fig. 1. Study area location and spatial distribution of the monitoring sites along Santiago River (above), and pollution hotspots along its topographic profile (below).

the Lerma River (among other agricultural, urban, and industrial wastewater discharges). Water management and pollution issues involving the Lerma and Santiago rivers, as well as Lake Chapala, have been studied previously and similar conclusions have been reached, concerning water quality and scarcity problems (Von Bertrab, 2003; Fall et al., 2007; Sedeño-Díaz and López-López, 2007; Cifuentes et al., 2011).

The foremost urban wastewater discharges to the Santiago River come from Guadalajara city, located in headwaters (Fig. 1). Two water pollution hotspots have been identified: northeast of the metropolitan area, near the so-called sector "*Agua Prieta*", and south of the urban zone, via "*El Ahogado*" stream (Rizo-Decelis and Andreo, 2016). Moreover, given the deteriorated water quality of the lake, a third pollution hotspot is considered. It is located near the first sampling site, northeast of Chapala Lake, by the source of the Santiago River, downstream from Ocotlán city.

2.2. Water quality data availability

The Water Commission of Jalisco State (CEA, 2016) provided the water quality datasets. The CEA conducts monthly monitoring of the water quality along the Santiago River, at ten sampling stations on the upper 262.5 km stretch, from north of Chapala Lake to La Yesca hydropower reservoir (1 to 10 in Fig. 1). The available datasets contained the analytical results of sampling campaigns carried out in the dry season (early October–late May), during four years (2009 to 2013). The value considered is the mean of those 32 measurements (8 months of the dry season, over 4 years), however, sometimes it may be some month that was not collected or an outlier attributed to inadequate collection. The latter was detected in a first screening of the data, then the used value is the mean of the considered measurements. For all the variables considered, the number of outliers that was omitted is <3%.

Moreover, during data processing, the values reported as blanks or not-detected (<6%) were replaced by half the value of the detection limit, depending on the equipment and chemical method, according to Croghan and Egeghy (2003).

There is a clear distinction between dry and wet seasons. According to Rizo-Decelis and Andreo (2016), the most representative hydrochemical data are those sampled during the dry season, when most of the variables concentration increase (Except for O_2) and the "statistical noise" (i.e. unexplained variations in samples) effect caused by the rainfall decreases.

A total of 28 variables were analyzed: water temperature (T), pH, dissolved oxygen (O_2), total coliform bacteria (TC), fecal coliform bacteria (FC), electric conductivity (EC), total dissolved solids (TDS), phosphorus (P), fluoride (F^-), methylene blue active substance (MBAS, i.e. surfactants), fats and oils (F&O), turbidity, total suspended soils (TSS), sedimentable solids (SS), alkalinity, hardness, nitrates (NO_3^-), ammoniac nitrogen (NH_3^+), nitrites (NO_2^-), total Kjeldahl nitrogen (NTK), iron (Fe), sodium (Na^+), zinc (Zn), chloride (Cl^-), sulfate (SO_4^{-2}), sulphide (S^-), biochemical oxygen demand (BOD_5), and chemical oxygen demand (COD).

The CEA laboratory techniques for analysis and sampling procedures are recognized by the International Laboratory Accreditation Cooperation (ILAC) and the International Accreditation Forum (IAF), through the Mexican Accreditation Entity (EMA). Additional details involving sampling collection, quality control procedures, and standards followed during campaigns, as well as laboratory analysis, are available on the website of the National Agency of Environment and Natural Resources of Mexico (SEMARNAT, 2010).

2.3. Definition of watershed support areas

The study area was subdivided into ten sub-basins, in view of the hydrographic-catchment extent of each sampling location. The support areas involving the upstream surface-water sampling sites were characterized by a single value for the variable (e.g. concentration), which was assumed to be representative along its sub-basin, since the sampling points represent the combination of the upstream values in the main watercourse and its tributaries.

The raster maps of the watersheds (or *support areas*) were discretized in a 5000 × 5000 m grid, in which every single sampling station was the ending discharge point for each sub-basin, based on the hydrological model obtained previously from a DEM (10 × 10 m), calculated from contour lines (INEGI, 2012), using ArcHydro 2.0 for ArcGIS 10.0 (ESRI, 1999–2010, 2013). Next, the support area was assigned with the observed value for that discharge point, according to the five-year monthly monitoring.

2.4. Geostatistical estimation methods

2.4.1. Ordinary kriging (OK)

According to Chiles and Delfiner (2012), kriging is the geostatistical estimator that has proven to be optimal for the spatial interpolation of environmental variables (Goovaerts, 1999; Webster and Oliver, 2007), including water quality. When the spatial locations of the observed data are scattered in space without physical constraints, on a plane (quality variables of groundwater measured from samples taken at monitoring wells) or in three-dimensional space (quality of atmospheric air measured anywhere on the Earth's surface), ordinary kriging (OK) is the standard optimal interpolator.

An important problem in environmental sciences is that given a set of n observed data $\{z(u_1),z(u_2),\ldots,z(u_n)\}$ of a given variable, there is an interest in estimating the value of the variable $z(.)$ at a non-sampled location u_0, namely the unknown value $z(u_0)$.

The OK estimate is given as a weighted average of the observation data:

$$Z^*_{OK}(u_0)=\sum_{i=1}^{n}\lambda_i Z(u_i) \tag{1}$$

where the optimal weights $\{\lambda(u_i),i=1,\ldots,n\}$ are obtained by solving the so-called OK system of equations (Olea, 1999):

$$\begin{aligned}&\sum_{i=1}^{n}\lambda_i\gamma_Z(u_\mathrm{i},u_\mathrm{j})+\mu_0=\gamma_Z(u_0,u_j)\\&j=1,\ldots,n\\&\sum_{i=1}^{n}\lambda_i=1\end{aligned} \tag{2}$$

where μ_0 is a Lagrange multiplier and $\gamma_Z(u_0,u_j)$ is the value of a semivariogram model of the raw variable and the Euclidean distance (u_0,u_j), which is between the j-th observation location and the location where the variable is going to be estimated.

OK also provides a measure of the uncertainty of the estimated value as given by the estimation variance:

$$\sigma^2_{OK}(u_0)=\sum_{i=1}^{n}\lambda_i\gamma_Z(u_0,u_i)+\mu_0 \tag{3}$$

OK can be improved in several ways. One way is by adding a secondary variable $\{D(u_i)\}$, which is linearly related with the expected value of the variable of interest:

$$\mathrm{E}\{Z(u_i)\}=a+bD(u_i) \tag{4}$$

2.4.2. Ordinary kriging with external drift (OKED)

The new estimator, proposed by Wackernagel (2003), is known as ordinary kriging with external drift (OKED):

$$Z^*_{OKED}(u_0)=\sum_{i=1}^{n}\lambda_i Z(u_i) \tag{5}$$

where the optimal weights $\{\lambda(u_i),i=1,\ldots,n\}$ are obtained by solving the so-called OKED system of equations (Hudson and Wackernagel, 1994; Wackernagel, 2003):

$$\begin{aligned}&\sum_{i=1}^{n}\lambda_i\gamma_R(u_i,u_j)+\mu_0+\mu_1D(u_j)=\gamma_R(u_0,u_j)\\&j=1,\ldots,n\\&\sum_{i=1}^{n}\lambda_i=1\\&\sum_{i=1}^{n}\lambda_iD(u_i)=D(u_0)\end{aligned} \tag{6}$$

Here, $\gamma_R(u_i,u_j)$ is the value of a semivariogram model of the residual variable and the Euclidean distance between u_i and u_j; besides, μ_0 and μ_1 are Lagrange multipliers. The residual variable is the difference between the original variable and the linear trend expressed in Eq. (4). Additionally, the estimation variance is given by:

$$\sigma^2_{OKED}(u_0)=\sum_{i=1}^{n}\lambda_i\gamma_R(u_i,u_j)+\mu_0+\mu_1D(u_0) \tag{7}$$

2.4.3. Regression kriging (RK)

If only the secondary information is used, one might surmise that the spatial interpolation could be obtained through the regression estimate:

$$Z^*_R(A_0)=\hat{a}+\hat{b}\tilde{D}(u_0) \tag{8}$$

with estimation variance:

$$\sigma_R^2(A_0) = \sigma_{Res}^2\left(1 + X_0^T\left(X^TX\right)^{-1}X_0\right) \tag{9}$$

where σ_{Res}^2 is the variance of the regression residuals, X is the *nx2* matrix of basis functions $\{(1, D\sim(u_i); i = 1, \ldots n)\}$ of the experimental data and X_0 is the 1×2 vector of the basis at the location to be estimated $\{1, D\sim(u_0)\}$. The superscript *T* represents the transpose of the matrix or vector. The residuals are given by:

$$R_R^*(u_i) = Z(A_i) - \hat{a} - \hat{b}\tilde{D}(u_i) \tag{10}$$

However, a more interesting estimator, known as regression kriging (RK), can be obtained if the regression residual at the unknown location is estimated by means of ordinary kriging (Hengl et al., 2007):

$$R_{OK}^*(u_0) = \sum_{i=1}^{n} \lambda_i R_R^*(u_i) \tag{11}$$

In addition, the estimate is added to the regression estimate to derive the RK estimate:

$$Z_{RK}^*(A_0) = Z_R^*(A_0) + R_{OK}^*(A_0) \tag{12}$$

The estimation variance by RK (σ_{RK}^2) is given by the addition of the variances of regression (σ_{Res}^2) of ordinary kriging residuals (σ_{OK}^2), since they are independent:

$$\sigma_{RK}^2(A_0) = \sigma_{Res}^2(A_0) + \sigma_{OK}^2(A_0) \tag{13}$$

That is:

$$\sigma_{RK}^2(u_0) = \sigma_{Res}^2\left(1 + X_0^T\left(X^TX\right)^{-1}X_0\right) + \sum_{i=1}^{n} \lambda_i \gamma_{Res}(u_0, u_j) + \mu_0 \tag{14}$$

2.4.4. Topological kriging

Interpolation of environmental variables represents another significant problem to be faced by means of geostatistical-based advanced solutions, as in the example of water quality variables when measured from streams or samples collected along a large river. In such cases, there is a restriction of physical locations where the value can be obtained in the field. Several possibilities could be considered in this case. For instance, one may take stream distances between observational measurements rather than Euclidean distances, as in Ver Hoef et al. (2006). Alternatively, the stream can be considered as having a support equal to its watershed (Fig. 2), as in Skøien et al. (2006), who put forth one of the first applications of kriging on stream networks. They defined TK as a *block-kriging* (Chiles and Delfiner, 2012), where the support of each observation location is its watershed. Downstream watersheds would contain the upstream watersheds, somewhat resembling the concept of a nesting support model (Fig. 3), which has been used with territorial units for water management statistics (Johnson, 2012) and other scientific studies (Mengistu et al., 2013). Therefore, the set of *n* observation data $\{A(u_1),$-$A(u_2), \ldots, A(u_n)\}$ of a given spatial environmental variable with watershed support is available, and the focus is on a value estimation of the variable of interest at a non-sampled watershed $A(u_0)$. Among the locations, number *1* represents a point while the remaining numbers (from 2 to 10) represent a watershed, as shown in Fig. 3.

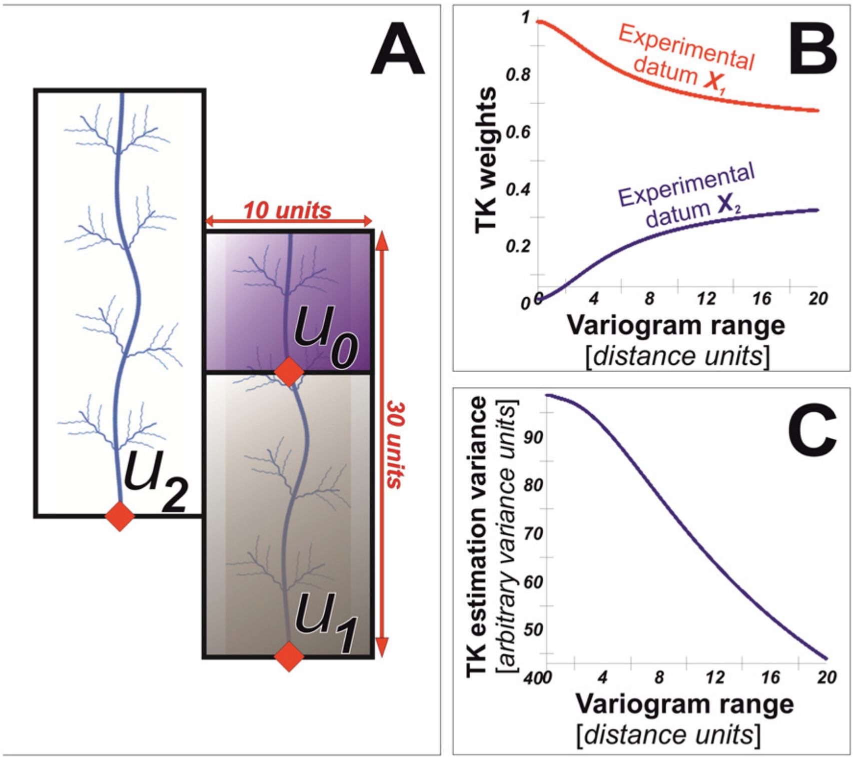


Fig. 2. Synthetic example of Topological kriging (based on Skøien et al., 2006). A: Layout of the problem where a watershed u_0 (in purple) is estimated using two observation data: watershed u_1 (in gray) that contains the basin to be estimated and watershed u_2, which does not overlap with u_0 or u_1. B: Weights assigned, by topological kriging, to each of the experimental data as a function of the range of the semivariogram of the underlying point process. The watershed that encloses the one to be estimated always gets more weight than the other observation datum. C: Topological kriging estimation variance as a function of the semivariogram range.

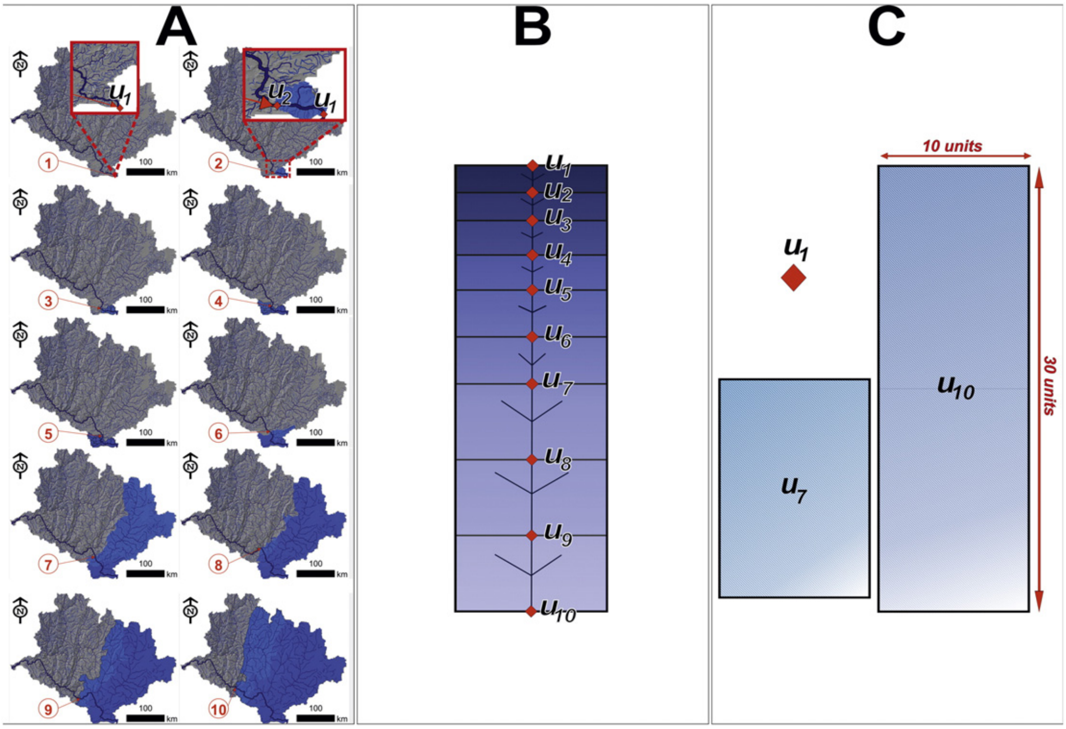


Fig. 3. A: Nested support model for the Santiago River study area. Only the first measurement ("1", next to Chapala Lake) has a point support. The rest of the observation data possess areal support (watershed), where each basin encloses the previous ones. B: Synthetic example that mimics A. C: Three experimental locations (watersheds) chosen to be estimated: 1 (the uppermost one, with a point support), 7 (a middle basin enclosing basins, 6, 5, 4, 3, 2, and point 1), and 10 (the lowest watershed, containing all other watersheds as well as point 1).

The estimated value by TK is given as a weighted average of the experimental data (Skøien et al., 2006):

$$Z^*_{TK}(A_0)=\sum_{i=1}^{n}\lambda_i Z(A_i) \tag{15}$$

where the optimal weights $\{\lambda(A_i), i=1,\ldots,n\}$ are obtained by solving the so-called TK system of equations (Skøien et al., 2006):

$$\begin{aligned}&\sum_{i=1}^{n}\lambda_i\gamma_{TK}(A_i,A_j)+\mu_0=\gamma_{TK}(A_0,A_j)\\ &j=1,\ldots,n\\ &\sum_{i=1}^{n}\lambda_i=1\end{aligned} \tag{16}$$

Here, $\gamma_{TK}(A_i,A_j)$ is the value of a semivariogram model of the variable with watershed support for the watersheds A_i and A_j.

The variance of estimation by TK can be written as:

$$\sigma^2_{TK}(A_0)=\sum_{i=1}^{n}\lambda_i\gamma_{TK}(A_0,A_i)+\mu_0 \tag{17}$$

2.4.5. Topological kriging with external drift (TKED)

TK can also be improved by considering a secondary variable $\{D\sim(u_i)\}$, which is linearly related to the variable of interest:

$$\mathrm{E}\{Z(A_i)\}=a+b\tilde{D}(u_i) \tag{18}$$

where a and b are the intercept with the origin and slope of the regression line, which is estimated from the experimental data pairs $\{(Z(A_i),D\sim(u_i)),i=1,\ldots,n\}$. In addition, the new notation $D\sim(u_i)$ is introduced for the secondary variable, which will be stream distance (in particular, measured to the upstream contamination hotspot).

The value estimated by TK with external drift (TKED) can be written as:

$$Z^*_{TKED}(A_0)=\sum_{i=1}^{n}\lambda_i Z(A_i) \tag{19}$$

The optimal values for the weights are obtained by solving the system of TKED (Laaha et al., 2013), as:

$$\begin{aligned}&\sum_{i=1}^{n}\lambda_i\gamma_R(A_i,A_j)+\mu_0+\mu_1\tilde{D}(u_j)=\gamma_R(A_0,A_j)\\ &j=1,\ldots,n\\ &\sum_{i=1}^{n}\lambda_i=1\\ &\sum_{i=1}^{n}\lambda_i\tilde{D}(u_i)=\tilde{D}(u_0)\end{aligned} \tag{20}$$

In addition, the TKED variance is given by:

$$\sigma^2_{TKED}=\sum_{i=1}^{n}\lambda_i\gamma_R(A_0,A_i)+\mu_0+\mu_1\tilde{D}(u_0) \tag{21}$$

2.4.6. Regression topological kriging (RTK)

Furthermore, it could be assumed that the residual is a function of the watershed, in order to arrive at a regression by topological kriging (RTK), defined as:

$$Z^*_{RTK}(A_0)=Z^*_R(A_0)+R^*_{TK}(A_0) \tag{22}$$

When the residual is estimated by TK rather than by OK, the estimation variance of RTK will be $\sigma^2_{RTK}(A_0)$.In the case of TK, the estimation of the semivariogram of the data with watershed support is a challenging aspect. This can be done by using the inverse procedure and Eq. (23). A semivariogram model is injected in Eq. (23) (right-hand side) for each pair (A_i,A_j) and a theoretical value $\gamma_{TK}(A_i,A_j)$ is obtained (left-hand side). These values can be compared with the observed $\gamma^*_{TK}(A_i,A_j)$ using Eq. (26), once the point model that minimizes the objective function (OF) has been chosen.

$$\begin{aligned}\gamma_{TK}(A_i,A_j) &= \frac{1}{2}\mathrm{Var}(Z(A_i)-Z(A_j))\\ &= \frac{1}{A_iA_j}\int_{A_i}\int_{A_j}\gamma_p(u_i,u_j)du_idu_j\\ &-\frac{1}{2}\left[\frac{1}{A_i^2}\int_{A_i}\int_{A_i}\gamma_p(u_i,u_j)du_idu_j+\frac{1}{A_j^2}\int_{A_j}\int_{A_j}\gamma_p(u_i,u_j)du_idu_j\right]\end{aligned} \tag{23}$$

Here, $\gamma_{TK}(A_i,A_j)$ is the semivariogram between two experimental stream locations with catchment areas A_i and A_j respectively, and $\gamma_p(u_i,u_j)$ is the point-semivariogram between the point locations u_i and u_j (i.e. from points u_i,u_j to discretized catchment areas A_i,A_j, as in Fig. 3).

When experimental supports to be estimated have the same size (watershed support) but different from the experimental support (point support), the process obtaining the semivariogram with watershed support (or block support) is as well-known case of regularization (Journel and Huijbregts, 1978; Skøien et al., 2006). However, in this case, the regularization must be done numerically by the different support sizes in the experimental information and the supports to be estimated.

TK considers Euclidean distances to account for the effect of underlying continuous layers (e.g. soil, lithology, topography, or vegetation) whose connectivity is not stream-related, but may still influence the estimated values of stream-related variables (Skøien et al., 2006).

The semivariogram is calculated from its definition:

$$\begin{aligned}\gamma_{TK}(A_i,A_j) &= \gamma_{ij} = \frac{1}{2}\mathrm{Var}(Z(A_i)-Z(A_j))\\ &= \frac{1}{2}\mathrm{E}(Z(A_i)-Z(A_j))^2-\frac{1}{2}\mathrm{E}^2(Z(A_i)-Z(A_j))\\ &= \frac{1}{2}\mathrm{E}(Z(A_i)-Z(A_j))^2\end{aligned} \tag{24}$$

Since there is a small number of experimental data, the semivariogram cloud will be used; it is half the squared difference between each pair of observed data:

$$\gamma^*_{TK}(A_i,A_j) = \gamma^*_{ij} = \frac{1}{2}(Z(A_i)-Z(A_j))^2 \tag{25}$$

The semivariogram cloud has $n(n-1)/2$ observed data of semivariogram values. Yet in TK, where each datum has a block support equal to its watershed, fitting cannot be done directly but rather by using inverse modeling and Eq. (23): i.e. $\gamma_p(u_i,u_j)$, which gives a theoretical value of γ_{ij}, to be compared with the observed value γ^*_{ij}, for all of the data pairs, using the objective function (OF) and Eq. (16), as follows:

$$OF = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(\gamma^*_{ij}-\gamma_{ij}\right)^2 \tag{26}$$

The model that minimizes the OF is chosen as the best model. In this case, the theoretical model (γ_{ij}) is the estimated point semivariogram model.

The Fig. 4 shows the semivariogram cloud for electric conductivity, which exemplifies the latter: if the data are considered with a point support (i.e., ignoring that they have a watershed as is done in TK), a model can be fitted by least squares as seen in the figure. On the right appear the experimental areal semivariogram points (solid blue dots). A point semivariogram model (left) induces (by Eq. (23)) an areal semivariogram (open red triangles) that can be compared with the experimental one (right). The fitted model is the one that minimizes those distances. The distance (x-axis) represents the range for the areal semivariogram

For convenience, we used *FORTRAN* to obtain the results, but the equations provided in this paper can be implemented with any general programming language. Interpolation methods may be applied using a common computer package for Geostatistics, such as SGeMS (2009). Practitioners could even modify the programs provided by geostatistical libraries such as *GSLIB* (Deustch and Journel, 1992) to derive their own

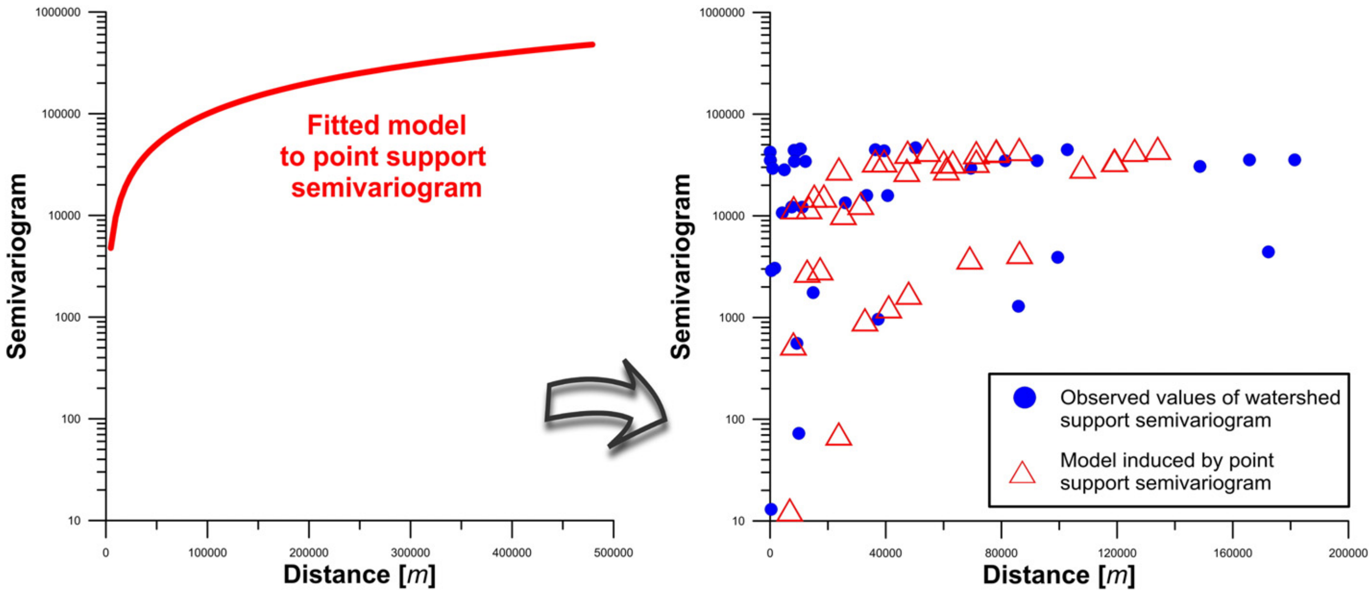


Fig. 4. Example of model fitted to the semivariogram cloud as an inverse model, applied to electric conductivity (EC) measured along the Santiago River waters. The task is to find a point support semivariogram model (red model on the left) that generates watershed support semivariogram values (red triangles in the right) that provides the best fit to the experimental watershed support semivariogram cloud values (blue circles in the right).

programs. Another possibility is to use an *R* package like "*Package rtop*" (Skøien et al., 2014; Skøien, 2015), which performs topological kriging.

Fig. 5 summarizes the screening criteria for the main interpolation characteristics of the prediction methods described above, according to the support variable criteria corresponding to each estimation model. It also condenses the systematic procedure behind the current methodological approach, in order to clarify the steps followed. An additional summary in simplified form (as table) is also included as Appendix A.

2.5. Cross-validation

The cross-validation technique, also known as the *Leave-One-Out* method, was used (Stone, 1974; Bradley, 1983) to compare the prediction results for each of the proposed estimators (OK, OKED, RK, TK, TKED, and RTK). This procedure omits one observed datum and predicts the concentration considering the rest of the observed data. The similarity between predictions and observed values indicates each model's performance. It was repeated for each observation point.

To soundly assess the accuracy of the predicted results from cross-validation, a number of statistics can be calculated for each of the evaluated methods (Webster and Oliver, 2007), such as Mean Error (*ME*) or Mean Square Normalized Error (*MSNE*).

$$ME(OK) = \frac{1}{n}\sum_{i=1}^{n} Z^{*}_{OK}(u_i) - Z(u_i) \tag{27}$$

$$MSNE(OK) = \frac{1}{n}\sum_{i=1}^{n} \frac{\left(Z^{*}_{OK}(u_i) - Z(u_i)\right)^2}{\sigma^2_{OK}(u_i)} \tag{28}$$

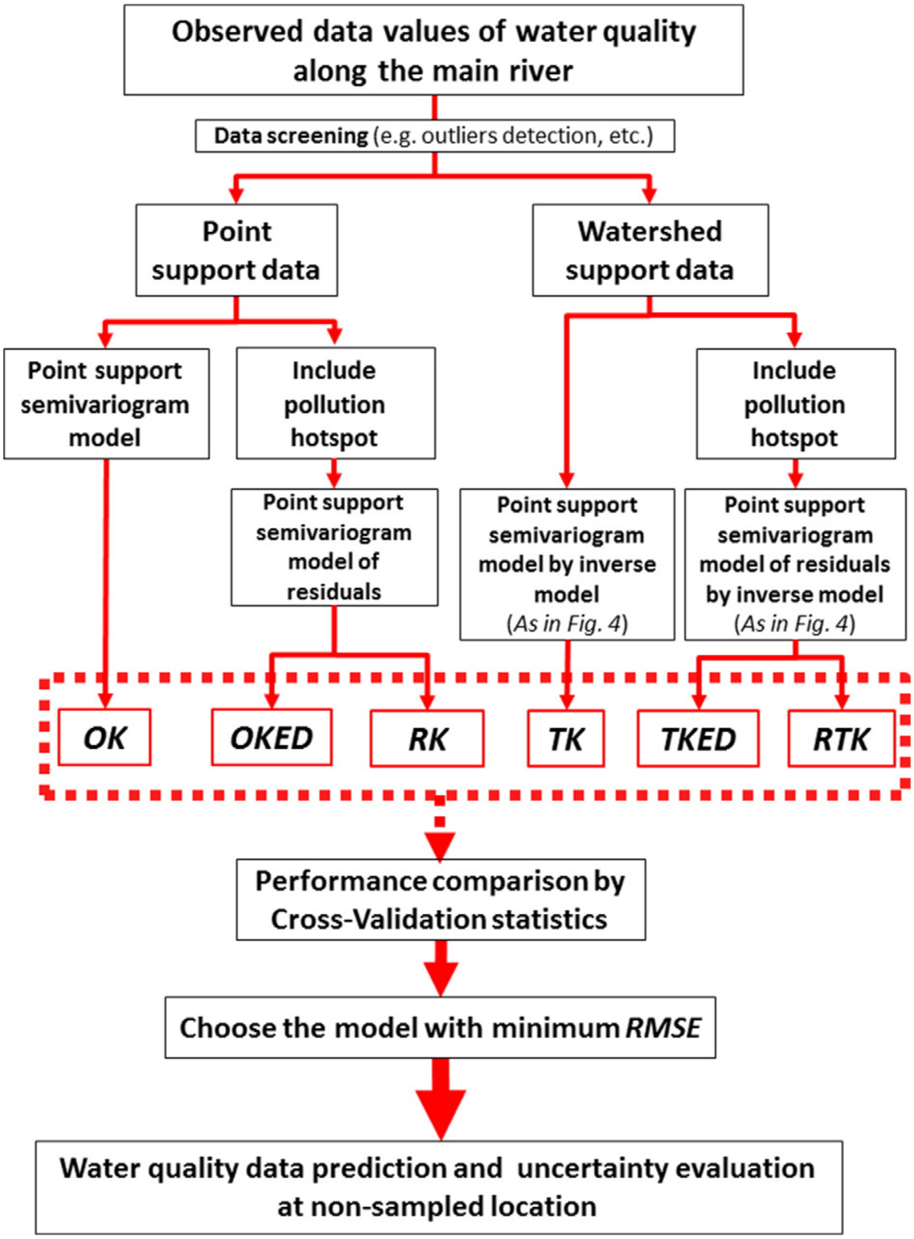

Observed data values of water quality along the main river
Data screening (e.g. outliers detection, etc.)
Point support data
Watershed support data
Point support semivariogram model
Include pollution hotspot
Include pollution hotspot
Point support semivariogram model of residuals
Point support semivariogram model by inverse model (As in Fig. 4)
Point support semivariogram model of residuals by inverse model (As in Fig. 4)
OK
OKED
RK
TK
TKED
RTK
Performance comparison by Cross-Validation statistics
Choose the model with minimum RMSE
Water quality data prediction and uncertainty evaluation at non-sampled location

Fig. 5. Flowchart of the process followed to develop this research. (*OK* = Ordinary kriging, *OKED* = Ordinary kriging with external drift, *RK* = Regression kriging, *TK* = Topological kriging, *TKED* = Topological kriging with external drift, *RTK* = Regression topological kriging, *RMSE* = Root Mean Square Error).

The one held to be most suitable is the root mean square error (*RMSE*), according to Hyndman and Koehler (2006), which is defined as:

$$RMSE(OK) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z^{*}_{OK}(u_i) - Z(u_i)\right)^2} \qquad (29)$$

Although notations for Eqs. (27)–(29) are for OK, the same statistics can be calculated for all methods. *ME* should be zero for an unbiased estimator (any form of kriging is unbiased by construction), and *MSNE* should be close to one if the evaluation of the uncertainty by the estimation variance is correct. The *RMSE* is always positive, and in general, the estimator with the lowest *RMSE* is preferable.

In practice, the cross-validation results can be used to estimate the semivariogram parameters. Thus the range is estimated as the value that minimizes the *RMSE*; and the variance (which has no influence on the *RMSE*).

The best *MSNE* value is equal to one (Hyndman and Koehler, 2006.). Hence,

$$MSNE(TK) = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(Z^{*}_{TK}(u_i) - Z(u_i)\right)^2}{\sigma^{*2}_{TK}}\sum_{i=1}^{n}\frac{\left(Z^{*}_{TK}(u_i) - Z(u_i)\right)^2}{\sigma^{*2}_{TK}\breve{\sigma}^{2}_{TK}(u_i)} = 1 \qquad (30)$$

and

$$\sigma^{*2}_{TK} = \sum_{i=1}^{n}\frac{\left(Z^{*}_{TK}(u_i) - Z(u_i)\right)^2}{\breve{\sigma}^{2}_{TK}(u_i)} \qquad (31)$$

where σ^{*2}_{TK} is the variance, whereas $\sigma^{2}_{TK}(u_i)$ is the estimation variance related with $Z^{*}_{TK}(u_i)$ by using a unit variance and an estimated range in order to minimize the *RMSE*.

3. Results and discussion

3.1. Cross-validation results

The charts shown in Figs. 6A and 6B display the cross-validation estimation results for selected water-quality variables, along with the predicted values for each of the seven methods considered (i.e., simple mean, RK, OK and OKED, TK, TKED, and RTK), in accordance with their respective measure units. The mean observed values are given for each of the 10 sampling stations along the Santiago River during the dry seasons, from 2009 to 2013.

Moreover, Fig. 7 summarizes the comparison between the *RMSE* calculated for all variables, using a *spot-light* color range, highlighting the lowest *RMSE* value in bold type.

The results of the *MSNE* cross-validation statistic can be consulted in Appendix B. The best *MSNE* value is equal to one. For the kriging methods, the *MSNE* is exactly one by construction, because the variance of the semivariogram is estimated in such a way (see Eqs. (30) and (31). On the other hand, for the regression methods the value is smaller than one. However, the latter results are not bad because it is in the safe side, with an estimated error variance larger than the true error variance.

3.2. Spatial prediction

Finally, each water quality parameter was predicted at 10 km-interval segments along the entire Santiago River profile. Confidence bounds of the estimation are included in order to identify the main gaps in information, considering the assessed error range of the predictions, which would also help optimize the monitoring network for water quality parameter values along the main watercourse of the basin. Four predicted variables (EC, P^{+}, COD, and Cl^{-}) are shown in Fig. 8. The dashed lines parallel (±) to the water-quality estimation curves (for methods TK, TKED, and RTK) indicate the confidence range defined for the prediction location, according to the square root of the obtained variance (i.e. ± the *RMSE*). When values below zero were obtained through the interpolation process, the lower limit of the confidence bounds was defined as zero, as it is meaningless to present estimated pollution concentration by means of negative values. In the case of COD, due to the logarithmic scale chosen in *y*-axes, confidence bounds defined as zero appear as blanks.

Additionally, Appendix C contains an example of the final map of P concentration, which has been estimated by RTK method, along prediction points of the Santiago River.

The cross-validation statistical error calculations (Fig. 7) indicate that the TK approach, and its combination with regression kriging (RTK), by far surpass the other approaches. TK and RTK are closer to the observed values than the rest of the prediction methodologies for most (79%) of the variables assessed in this paper, with 39% (TK), 36% (RTK) and 4% (TKED) of the total, respectively. Accordingly, TK and RTK are positively the best predicting methods for wastewater discharge occurrences (both industrial and urban wastewaters). This is clearest for the most common variables (COD, BOD_5, P, EC, Cl^{-}. NH_3^{-}, MBAS, TDS, SO_4^{-2}, SS, TSS, turbidity, alkalinity, Fe, Zn, or F^{-}) associated with water contamination, particularly with regard to nitrogen compounds, phosphates, and oxygen demand, the leading water quality variables monitored in rivers worldwide. In some other outstanding cases, such as hardness, TK is significantly (five times) better than OK (Fig. 7).

In certain cases, such as the variables MBAS, TDS, O_2, EC, BOD_5, COD, and Nitrogen and Sulphide compounds, TK tends to underestimate the values at the station located in the lower basin area, around La Yesca hydropower dam (sampling station No. 10, in Fig. 8). This may be due to the relatively great size of the support area for that sampling station. The sampling points located downstream carry larger sub-basin surface (and catchment) areas, which implies a greater dilution effect of the measured variables. The rainfall increasing in the lower part of the watershed might also contribute to this dilution effect. Although overall the predictions are more accurate using TK and RTK, both O_2 and S^{-} (which are two important variables for describing water quality in rivers) seem to be better estimated by simple RK than via the other estimators. As affirmed by Chang (2008), spatial regression models are more accurate in explaining water quality variations than other models, a fact corroborated by the current results.

In the specific case of O_2 and sulfides, the RK surpasses OK in accuracy, and its linear regression was not strongly correlated with the upstream hotspot location.

Some other water quality variables (e.g. FC, TC, pH, Fe) are estimated more accurately by RK and OK, rather than by TK, TKED, or RTK (about 18% of the total). This might be attributed to the lack of correlation between watershed area changes along the river channel, and concentration changing for these variables.

Similarly, T is ~80% more accurate when estimated by RTK as opposed to OK, accomplishing a precision in prediction of 0.74 °C (Fig. 8).

It is remarkable that concentrations of coliform bacteria (*FC* and *TC*) show important changes along the Santiago River. The fact that differences of several orders of magnitude were detected between sampling stations for *TC* and *FC* makes it even more difficult to estimate them properly by means of any prediction interpolation model (Fig. 7). There is no apparent association with the pollution hotspots (as an external drift or spatial regression), and/or the expected dilution due to the watershed catchment area increasing downstream (topologically). The best interpolation method for these variables seems to be RK, despite a considerably high value for *RMSE*, many times the value of the maximum permissible limit for in river waters around most of the world (e.g. ±1000 MPN/100 ml).

The basin-size difference, in terms of supporting area dimensions between sampling stations 6 and 7 (Fig. 3), could be linked to dilution processes making water quality prediction by TK very inaccurate (Fig.

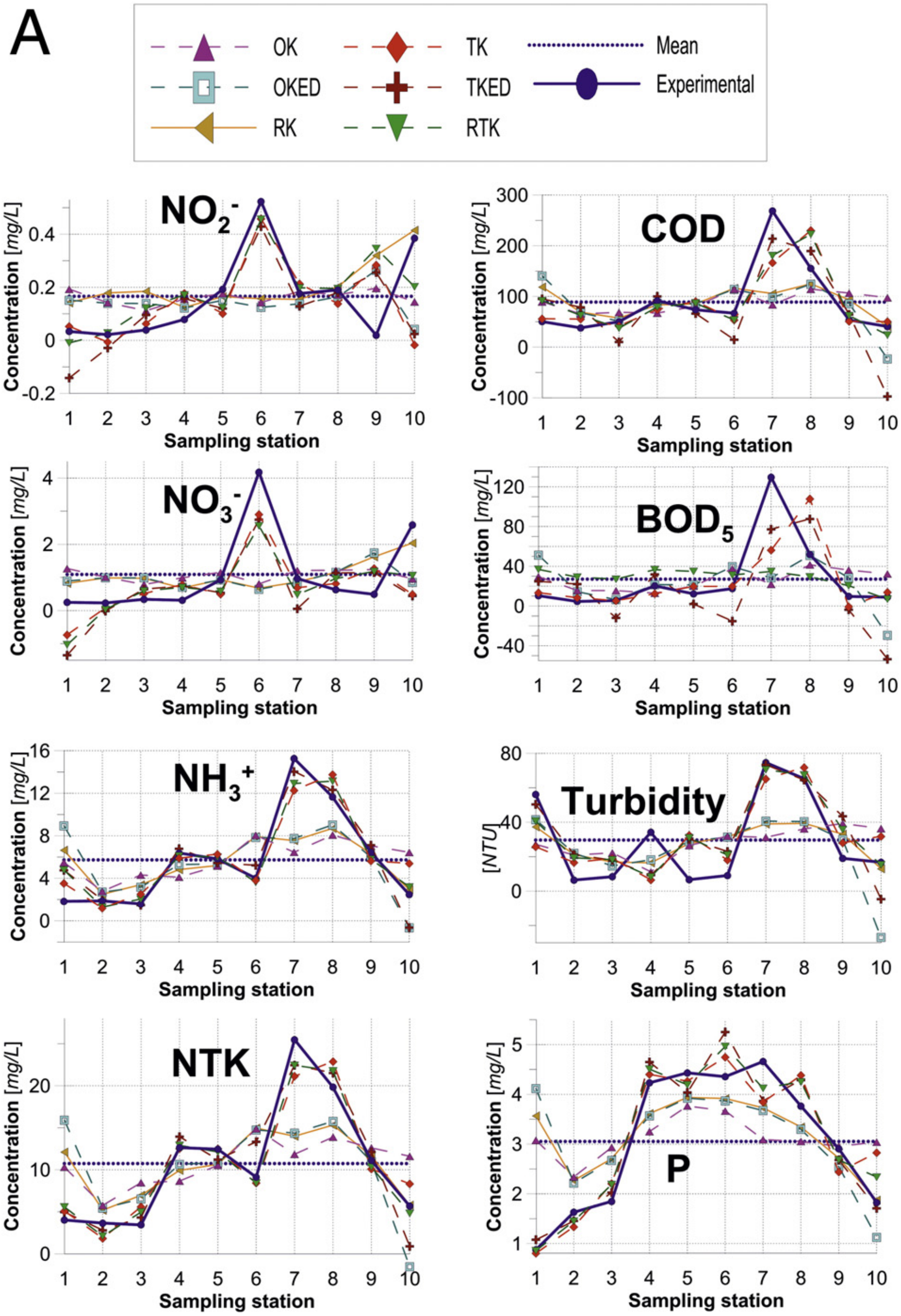


Fig. 6A. Cross-validation results, for each of the six estimation methods assessed, for the variables related to wastewater discharges. (OK = Ordinary kriging, RK = Regression kriging, OKED = Ordinary kriging with external drift, TK = Topological kriging, TKED = Topological kriging with external drift, RTK = Regression topological kriging).

6A and B), most of the assessed models estimation failed at that segment of the river.

The pH is the only variable predicted better by OK along the Santiago River, and it shows only slight variations. In the case of the variable indicative of "fat and oils" (*F&O*) content, that is related to urban and industrial process wastewaters (Williams et al., 2012), it reflects no correlation with the identified pollution hotspots, nor with the watershed catchment area increasing downstream. Because the behavior of *F&O* concentrations is erratic, the best estimator would be the simple mean (Fig. 7).

According to the above results, enhanced water quality assessment along the Santiago River would call for improved monitoring, maybe by relocating some sampling stations and densifying the monitoring (spatially and temporally). A lower number of sampling stations could be installed in the lower zone, given that confidence ranges show a relatively smaller uncertainty in this area (Fig. 8). On the other hand, for most of the predicted variables of the watershed, in the upper zone (close to Chapala Lake and downstream El Ahogado - Fig. 1) there is a greater need of increasing the data collected in situ. The latter is supported by the fact that stations *6* and *7* register a significant increase in chemical oxygen demand concentrations.

Efforts were made to improve the performance of the proposed approach by using the pollution hotspots as external drift. However, it was found that the pollution hotspots location only improve predictions if it is considered in the regression model. Although the *Agua Prieta* zone location has been considered as a wastewater pollution hotspot (Fig. 1), as well as the *El Ahogado* stream confluence with the Santiago River, it is important to note that there are further punctual discharge sources in

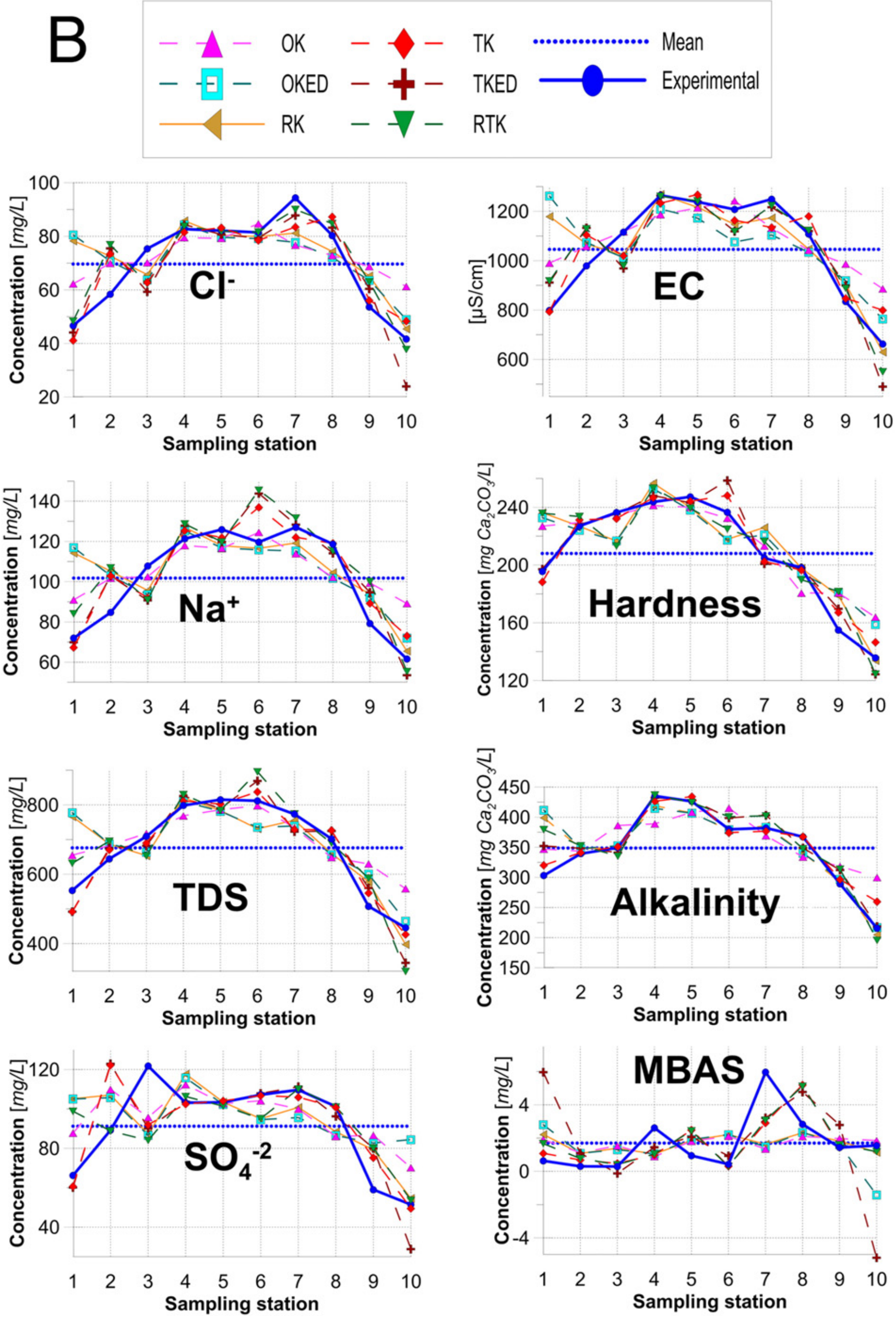
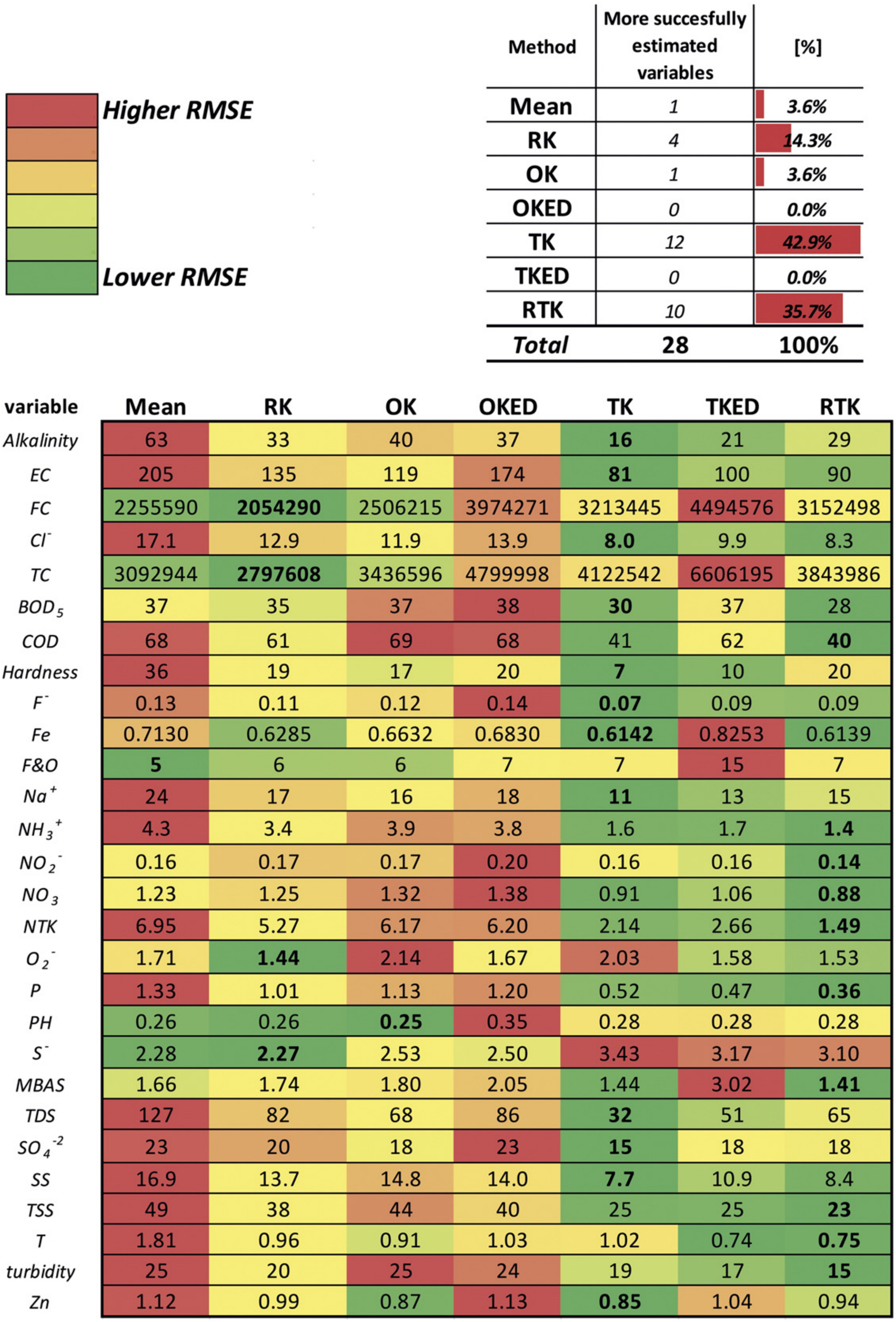


Fig. 6B. Cross-validation results, for each of the six estimation methods assessed, for the variables related to water mineralization processes. (OK = Ordinary kriging, RK = Regression kriging, OKED = Ordinary kriging with external drift, TK = Topological kriging, TKED = Topological kriging with external drift, RTK = Regression topological kriging).

the surrounding area, for instance in the northern zone of Guadalajara city. Since 2012 the Mexican government has built two wastewater treatment plants (WTP), located precisely downstream from the *El Ahogado* and the *Agua Prieta* streams. These WTP have capacity to treat 2.3 m^3/s and 8.5 m^3/s, respectively. Yet until of December 2013, there were no clear signs of water pollution diminishing for most of the indicator parameters downstream (except for MBAS, *F&O* and turbidity). Possibly, a certain part of a given urban wastewaters discharge flow is not entering to the primary sewer WTP. It possible represents an indicative of diffuse pollution as second contamination cause, which have not been considered in this first approach, and deserves further research.

As samples in the current monitoring are taken at the main river channel exclusively, the dataset would not seem to be the most appropriate, whereas TK is most advantageous for dendritic structures with neighboring catchments. Yet, in the current study, upstream catchments are given less weight mainly because of their smaller size and larger distance to the downstream sections, which is similar to OK. On the other hand, results show that TK outperforms OK for this dataset, which underlines the benefits of TK even for such restricted data. Therefore, predictions might improve substantially if data from tributaries were included, and it would illustrate more clearly the advantages of TK method approach. It definitely deserves further research.

The prediction locations represent the mean water quality estimation for a specific variable, taking into consideration its basin as a supporting area (Fig. 8). Therefore, the predicted value is only valid and representative at that one particular stretch along the Santiago River.


Higher RMSE
Lower RMSE

Method	More succesfully estimated variables	[%]
Mean	*1*	*3.6%*
RK	*4*	*14.3%*
OK	*1*	*3.6%*
OKED	*0*	*0.0%*
TK	*12*	*42.9%*
TKED	*0*	*0.0%*
RTK	*10*	*35.7%*
Total	**28**	**100%**

variable	Mean	RK	OK	OKED	TK	TKED	RTK
Alkalinity	63	33	40	37	**16**	21	29
EC	205	135	119	174	**81**	100	90
FC	2255590	**2054290**	2506215	3974271	3213445	4494576	3152498
Cl^-	17.1	12.9	11.9	13.9	**8.0**	9.9	8.3
TC	3092944	**2797608**	3436596	4799998	4122542	6606195	3843986
BOD_5	37	35	37	38	**30**	37	28
COD	68	61	69	68	41	62	**40**
Hardness	36	19	17	20	**7**	10	20
F^-	0.13	0.11	0.12	0.14	**0.07**	0.09	0.09
Fe	0.7130	0.6285	0.6632	0.6830	**0.6142**	0.8253	0.6139
F&O	**5**	6	6	7	7	15	7
Na^+	24	17	16	18	**11**	13	15
NH_3^+	4.3	3.4	3.9	3.8	1.6	1.7	**1.4**
NO_2^-	0.16	0.17	0.17	0.20	0.16	0.16	**0.14**
NO_3	1.23	1.25	1.32	1.38	0.91	1.06	**0.88**
NTK	6.95	5.27	6.17	6.20	2.14	2.66	**1.49**
O_2^-	1.71	**1.44**	2.14	1.67	2.03	1.58	1.53
P	1.33	1.01	1.13	1.20	0.52	0.47	**0.36**
PH	0.26	0.26	**0.25**	0.35	0.28	0.28	0.28
S^-	2.28	**2.27**	2.53	2.50	3.43	3.17	3.10
MBAS	1.66	1.74	1.80	2.05	1.44	3.02	**1.41**
TDS	127	82	68	86	**32**	51	65
SO_4^{-2}	23	20	18	23	**15**	18	18
SS	16.9	13.7	14.8	14.0	**7.7**	10.9	8.4
TSS	49	38	44	40	25	25	**23**
T	1.81	0.96	0.91	1.03	1.02	0.74	**0.75**
turbidity	25	20	25	24	19	17	**15**
Zn	1.12	0.99	0.87	1.13	**0.85**	1.04	0.94

Fig. 7. Comparison of statistics for the calculated *RMSE* of the 28 selected water quality variables, derived from Cross-Validation results (RK = Regression kriging; OK = Ordinary kriging; OKED = Ordinary kriging with external drift, TK = Topological kriging, TKED = Topological kriging with external drift, RTK = Regression topological kriging).

4. Conclusions

In this paper, a methodological approach for water quality estimation in rivers has been introduced and tested. According to the cross-validation results, the Top-kriging (TK) method offers a more accurate water quality prediction than the others (including Ordinary Kriging) for many of the measured parameters along the Santiago River, in Mexico. Most remarkable are the cases of quality variables closely

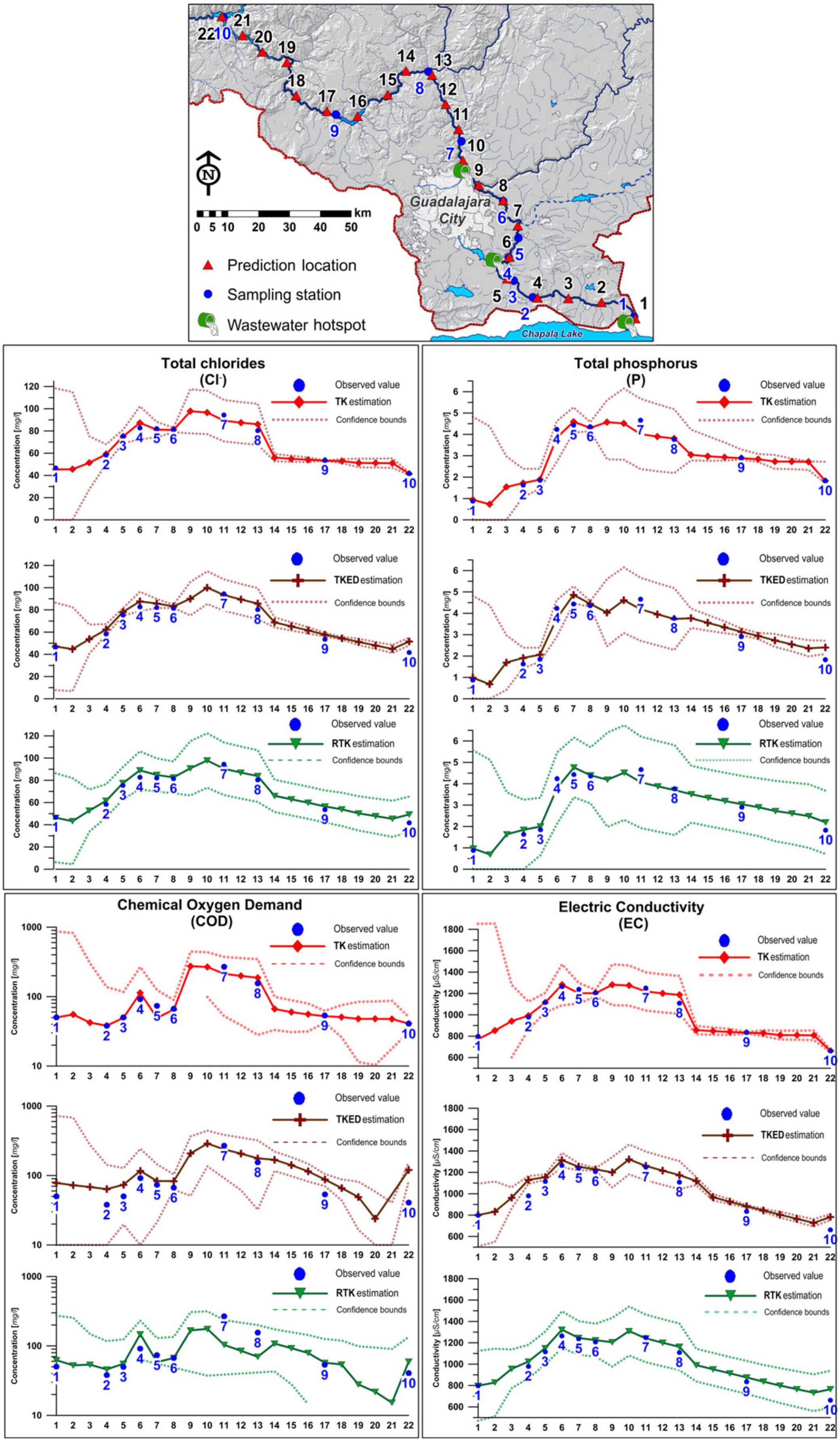


Fig. 8. Spatial prediction of four selected water quality variables, at 10 km-interval segments along the Santiago River profile.

involved with wastewater discharge. Consequently, spatial regression models, addressing the nested basin and its surface area, explain more adequately water quality dynamics.

In order to predict water quality in large rivers, the TK, and its combination with regression estimators (i.e. RTK), shows the highest efficiency for estimation of many variables, achieving higher accuracy than OK and greater precision than simple regression, OKED, or the simple mean. Notwithstanding, one size does not fit all: no single method for water quality prediction can be held up as the best in all settings. Key aspects to be taken into account are the natural characteristics of the studied area, the variables of interest, and the available monitoring data.

There is a need to properly process available data, and to fully exploit it by means of a reasonable application of adequate interpolation methods. The persistent goals are to improve monitoring and water quality assessment, to infer the water quality in non-sampled sites, and most importantly, to refine monitoring strategies. The approach presented here can be applied effectively in the case of large river systems where detailed topographic information is available, and where the foremost localized wastewater discharge has been previously identified; such is the case of the Santiago River in Mexico.

Incorporating watershed extension and spatial regression throws new light on the quality variable dynamics of the Santiago River. Future water quality management in the Santiago River Watershed should therefore accommodate recommendations for optimizing water quality monitoring, by considering the spatial correlations between pollution hotspots location and sub-basin areas of the sampling points, in addition to regarding the key quality parameters. The contamination hotspot location, however, does not appear to be significant for most variables.

Acknowledgements

The current research is supported by a PhD grant (#312640) from the Consejo Nacional de Ciencia y Tecnología (CONACYT) of Mexico, assigned to the first author. The authors kindly thank the four anonymous reviewers for providing constructive criticism. We are grateful to Julie Lavoie and Jean Sanders for improving the English text. This article is a contribution to the project CGL2015-65858-R of DGICYT, and to the Research Group RNM-308 of the Andalusian Government.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.scitotenv.2017.06.145.

References

Álvarez-Cabria, M., Barquín, J., Peñas, F.J., 2016. Modelling the spatial and seasonal variability of water quality for entire river networks: relationships with natural and anthropogenic factors. Sci. Total Environ. 545-546:152–162. http://dx.doi.org/10.1016/j.scitotenv.2015.12.109.

Ani, E., Hutchins, M., Kraslawski, C., Agachi, P., 2011. Mathematical model to identify nitrogen variability in large rivers. River Res. Appl. 27 (10):1216–1236. http://dx.doi.org/10.1002/rra.1418.

Bradley, E., 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. J. Am. Stat. Assoc. 78 (382):316–331. http://dx.doi.org/10.1080/01621459.1983.10477973.

Brammer, S., 2014. Domaining bi-modal data sets geostatistically using a directional neighborhood search. In: Pardo-Igúzquiza, E., Guardiola-Albert, C., Heredia, J., Moreno-Merino, L., Durán, J., Vargas-Guzmán, J. (Eds.), Mathematics of Planet Earth. Lecture Notes in Earth System Sciences. Springer, Berlin, Heidelberg http://dx.doi.org/10.1007/978-3-642-32408-6_168.

CEA, 2016. Water Quality Monitoring Results Along the Santiago River in Jalisco State (In Spanish, Date of access 04/01/2016, http://info.ceajalisco.gob.mx/sca/).

Chang, H., 2008. Spatial analysis of water quality trends in the Han River basin, South Korea. Water Res. 42 (13):3285–3304. http://dx.doi.org/10.1016/j.watres.2008.04.006.

Chiles, J.P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. 2nd edition. John Wiley & Sons, Inc. http://dx.doi.org/10.1002/9781118136188 (ISBN: 978-0-470-18315-1, 734).

Cifuentes, E., Lozano, F., Trasande, L., Goldman, R.H., 2011. Resetting our priorities in environmental health: an example from the south–north partnership in Lake Chapala, Mexico. Environ. Res. 111 (6):877–880. http://dx.doi.org/10.1016/j.envres.2011.05.017.

CONAGUA, 2015. Strategic projects for drinking water, sewerage and sanitation, according to the National Infrastructure Program 2014–2018. (Date of access 04/01/2016). http://www.conagua.gob.mx/english07/publications/StrategicProjects.pdf.

Croghan, C.W., Egeghy, P.P., 2003. Methods of dealing with values below the limit of detection using SAS, USEPA research documents. (Date of access 04/01/2016). http://analytics.ncsu.edu/sesug/2003/SD08-Croghan.pdf.

De Anda, J., Quiñones-Cisneros, S.E., French, R.H., Guzmán, M., 1998. Hydrologic balance of Lake Chapala (Mexico). JAWRA J. Am. Water Resour. Assoc. 34 (6):1752-1688. http://dx.doi.org/10.1111/j.1752-1688.1998.tb05434.x.

De Anda, J., Shear, H., Maniak, U., Riedel, G., 2000. Phosphorus balance in Lake Chapala, lakes & reservoirs: research and management. J. Great Lakes Res. 26 (2):129–140. http://dx.doi.org/10.1016/S0380-1330(00)70680-0.

Environmental Systems Research Institute (ESRI), 1999–2010. ArcMap Version 10.0, GIS Software.

Environmental Systems Research Institute (ESRI), 2013. ArcHydro Tools 2.0 for Version 10.0 of ArcMap GIS Software.

Fall, C., Hinojosa-Peña, A., Carreño-De-León, M.C., 2007. Design of a monitoring network and assessment of the pollution on the Lerma River and its tributaries by wastewaters disposal. Sci. Total Environ. 373 (1):208–219. http://dx.doi.org/10.1016/j.scitotenv.2006.10.053.

Ferrari, L., Pasquarè, G., Venegas-Salgado, S., Romero-Ríos, F., 1999. Geology of the western Mexican Volcanic Belt and adjacent Sierra Madre Occidental and Jalisco block. Geol. Soc. Am. Spec. Pap. 334:65–83. http://dx.doi.org/10.1130/0-8137-2334-5.65.

Garreta, V., Monestiez, P., Ver Hoef, J.M., 2010. Spatial modelling and prediction on river networks: up model, down model or hybrid? Environmetrics 21 (5):439–456. http://dx.doi.org/10.1002/env.995.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Applied Geostatistics Series. Oxford University Press, p. 512 (ISBN-3:9780195115383),.

Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89 (1-2):1–45. http://dx.doi.org/10.1016/S0016-7061(98)00078-0.

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Comput. Geosci. 33 (10):1301–1315. http://dx.doi.org/10.1016/j.cageo.2007.05.001.

Herdendorf, C.E., 1982. Large lakes of the world. J. Great Lakes Res. 8 (3):379–412. http://dx.doi.org/10.1016/S0380-1330(82)71982-3.

Hudson, G., Wackernagel, H., 1994. Mapping temperature using kriging with external drift: theory and an example from Scotland. Int. J. Climatol. 14 (1):77–91. http://dx.doi.org/10.1002/joc.3370140107.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4):679–688. http://dx.doi.org/10.1016/j.ijforecast.2006.03.001.

Instituto Mexicano para el Desarrollo Comunitario A.C. (IMDEC), 2007. Report on violations to the right to health and to a safe environment in Juanacatlán and El Salto, Jalisco, Mexico. (Date of access 10/02/2016). http://www2.ohchr.org/english/issues/globalization/business/docs/ExecutiveSummarySantiagoRiver_en.pdf.

Instituto Nacional de Estadística y Geografía (INEGI), 2012. Topographic mapping from the Mexican National Institute for Statistics and Geography. (Date of access 10/02/2016). http://www.inegi.gob.mx.

Instituto Nacional de Estadística y Geografía (INEGI), 2015. Land use mapping (1: 250.000) from the Mexican National Institute for Statistics and Geography. (Date of access 04/29/2017). http://www.inegi.gob.mx.

Isaak, D.J., Peterson, E.E., Ver Hoef, J.M., Wenger, S.J., Falke, J.A., Torgersen, C.E., Sowder, C., Steel, E.A., Fortin, M.J., Jordan, C.E., Ruesch, A.S., Som, N., Monestiez, P., 2014. Applications of Spatial Statistical Network Models to Stream Data. 1(3). WIREs Water Wiley Periodicals Inc:pp. 277–294. http://dx.doi.org/10.1002/wat2.1023.

Johnson, C., 2012. Toward post-sovereign environmental governance? Politics, scale, and EU water framework directive. Water Altern. 5 (1), 83–97 (Date of access 10/02/2016, http://www.water-alternatives.org/index.php/alldoc/articles/vol5/v5issue1/159-a5-1-6/file).

Deutsch, C.V., Journel, A., 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, p. 340 (ISBN-13: 978-0195100150).

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, UK, p. 612 (ISBN-10: 1930665911),.

Laaha, G., Skøien, J.O., Blöschl, G., 2012. Quantitative Geology and Geostatistics, Comparing Geostatistical Models for River Networks. Geostatistics Oslo 2012. 17: pp. 543–553. http://dx.doi.org/10.1007/978-94-007-4153-9_44.

Laaha, G., Skøien, J.O., Nobilis, F., Blöschl, G., 2013. Spatial prediction of stream temperatures using top-kriging with an external drift. Environ. Model. Assess. 18 (6): 671–683. http://dx.doi.org/10.1007/s10666-013-9373-3.

Laaha, G., Skøien, J.O., Blöschl, G., 2014. Spatial prediction on river networks: comparison of top-kriging with regional regression. Hydrol. Process. 28 (2):315–324. http://dx.doi.org/10.1002/hyp.9578.

Mengistu, S.G., Creed, I.F., Kulperger, R.J., Quick, C.G., 2013. Russian nesting dolls effect – using wavelet analysis to reveal non-stationary and nested stationary signals in water yield from catchments on a northern forested landscape. Hydrol. Process. 27 (5):1099-1085. http://dx.doi.org/10.1002/hyp.9552.

Meyer, S., Blaschek, M., Duttmann, R., Ludwig, R., 2015. Improved hydrological model parametrization for climate change impact assessment under data scarcity — the potential of field monitoring techniques and geostatistics. Sci. Total Environ. 543 (B): 906–923. http://dx.doi.org/10.1016/j.scitotenv.2015.07.116.

Moore, G., Marone, C., Carmichael, I.S.E., Renne, P., 1994. Basaltic volcanism and extension near the intersection of the Sierra Madre volcanic province and the Mexican Volcanic

Belt. Geol. Soc. Am. Bull. 106 (3):383–394. http://dx.doi.org/10.1130/0016-7606(1994)106<0383:BVAENT>2.3.CO;2.

Morio, M., Finkel, M., Martac, E., 2010. Flow guided interpolation - a GIS-based method to represent contaminant concentration distributions in groundwater. Environ. Model. Softw. 25 (12):1769–1780. http://dx.doi.org/10.1016/j.envsoft.2010.05.018.

Müller, M.F., Thompson, S.E., 2015. TopREML: a topological restricted maximum likelihood approach to regionalize trended runoff signatures in stream networks. Hydrol. Earth Syst. Sci. (HESS) 19:2925–2942. http://dx.doi.org/10.5194/hessd-12-1355-2015.

Namour, P., Schmitt, L., Eschbach, D., Bertrand, M., Fantino, G., Bordes, C., Breil, P., 2015. Stream pollution concentration in riffle geomorphic units (Yzeron basin, France). Sci. Total Environ. 532:80–90. http://dx.doi.org/10.1016/j.scitotenv.2015.05.057.

Nel, J.L., Roux, D.J., Abell, R., Ashton, P.J., Cowling, R.M., Higgins, J.V., Thieme, M., Viers, J.H., 2009. Progress and challenges in freshwater conservation planning. Aquat. Conserv. Mar. Freshwat. Ecosyst. 19 (4):474–485. http://dx.doi.org/10.1002/aqc.1010.

Olea, R.A., 1999. Geostatistics for Engineers and Earth Scientists. 303. Springer, US. http://dx.doi.org/10.1007/978-1-4615-5001-3 (ISBN 978-1-4613-7271-4).

Pardo-Igúzquiza, E., 1998. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. J. Hydrol. 210 (1-4):206–220. http://dx.doi.org/10.1016/S0022-1694(98)00188-7.

Rizo-Decelis, L.D., Andreo, B., 2016. Water quality assessment of the Santiago River and attenuation capacity of pollutants downstream Guadalajara City, Mexico. River Res. Appl. 32 (7):1505–1516. http://dx.doi.org/10.1002/rra.2988.

Rojas-Ortuste, F., 2014. Policy and Institutional Framework for Drinking Water and Sanitation in Latin America and the Caribbean (in Spanish: "Políticas e institucionalidad en materia de agua potable y saneamiento en América Latina y el Caribe"). United Nations Natural Resources and Infrastructure Series (ISSN 1680-9017, Date of access 04/01/2016, http://repositorio.cepal.org/bitstream/handle/11362/36776/S2014277_es.pdf?sequence=1).

Sedeño-Díaz, J.E., López-López, E., 2007. Water quality in the Río Lerma, Mexico: an overview of the last quarter of the twentieth century. Water Resour. Manag. 21 (10): 1797–1812. http://dx.doi.org/10.1007/s11269-006-9128-x.

SEMARNAT, 2010. Mexican official standards. Mexico Ministry of Environment and Natural Resources (in Spanish). (Date of access 04/01/2016). http://www.semarnat.gob.mx/leyes-y-normas/nmx-agua.

Skøien, J.O., 2015. Interpolation of data with variable spatial support. Rtop-package: a package providing methods for analysis and spatial interpolation of data with an irregular support, version 0.5-5, repository CRAN. (Date of access 05/09/2016). http://cran.at.r-project.org/web/packages/rtop/rtop.pdf.

Skøien, J.O., Merz, R., Blöschl, G., 2006. Top-kriging – geostatistics on stream networks. Hydrol. Earth Syst. Sci. 10 (2):277–287. http://dx.doi.org/10.5194/hess-10-277-2006.

Skøien, J.O., Blöschl, G., Laaha, G., Pebesma, E., Parajka, J., Viglione, A., 2014. Rtop: an R package for interpolation of data with a variable spatial support, with an example from river networks. Comput. Geosci. 67:180–190. http://dx.doi.org/10.1016/j.cageo.2014.02.009.

SMN, 2015. National meteorological service of Mexico website (in Spanish). (Date of access 04/01/2016). http://smn.cna.gob.mx.

Stanford Geostatistical Modeling Software (SGeMS), 2009. (Date of access 05/09/2016). http://sgems.sourceforge.net.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. B (Methodol.) 36 (2), 111–147 (Stable URL: http://links.jstor.org/sici?sici=0035-9246%281974%2936%3A2%3C111%3ACCAAOS%3E2.0.CO%3B2-W).

Tsuzuki, Y., 2015. Relationships between pollutant discharge and water quality in the rivers from "better" to "worse" water quality. Ecol. Indic. 52:256–269. http://dx.doi.org/10.1016/j.ecolind.2014.12.001.

Ver Hoef, J.M., Peterson, E., Theobald, D., 2006. Spatial models that use flow and stream distance. Environ. Ecol. Stat. 13 (4):449–464. http://dx.doi.org/10.1007/s10651-006-0022-8.

Von Bertrab, E., 2003. Guadalajara's water crisis and the fate of Lake Chapala - a reflection of poor water management in Mexico. Environ. Urban. 15 (2):127–140. http://dx.doi.org/10.1177/095624780301500204.

Wackernagel, H., 2003. Multivariate Geostatistics - An Introduction with Applications. 3rd ed. XV. Springer-Verlag, Berlin:p. 388. http://dx.doi.org/10.1007/978-3-662-05294-5 (ISBN 978-3-662-05294-5),.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. 2nd ed. John Wiley & Sons, Inc. (ISBN 978-0-470-02858-2, 330).

Williams, J.B., Clarkson, C., Mant, C., Drinkwater, A., May, E., 2012. Fat, oil and grease deposits in sewers: characterisation of deposits and formation mechanisms. Water Res. 46 (19):6319–6328. http://dx.doi.org/10.1016/j.watres.2012.09.002.

Yang, X., Jin, W., 2010. GIS-based spatial regression and prediction of water quality in river networks: a case study in Iowa. J. Environ. Manag. 91 (10):1943–1951. http://dx.doi.org/10.1016/j.jenvman.2010.04.011.