Systems Research and Info. Systems
Vol. 8, pp. 257–272
Reprints available directly from the publisher
Photocopying permitted by license only

ROBUST SPATIAL DATA ANALYSIS OF LAKE GENEVA SEDIMENTS WITH S+SPATIALSTATS

REINHARD FURRER[a,*] and MARC G. GENTON[b]

[a]*Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland;* [b]*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA*

(Received 11 May 1998)

This paper discusses the use of robust geostatistical methods on a multivariate data set of sediments in Lake Geneva in Switzerland. Each variable is detrended via nonparametric estimation penalized with a smoothing parameter. The optimal trend is computed with a smoothing parameter based on cross-validation. Then, variograms are estimated by a highly robust estimator of scale. The parametric variogram models are fitted by generalized least squares, thus taking account of the variance–covariance structure of the variogram estimates. Kriging has been performed inside the Lake Geneva boundaries, and results are in close agreements with the geographical surroundings. The comparison of the kriging results with and without detrending the data relieved the importance of the trend detection and trend removing, and that a simple model with constant trend for this data set is not satisfactory. All these computations are done with the software S+SPATIALSTATS, extended with new functions in S+ that are made available.

Keywords: Robustness; trend; variogram; generalized least squares; kriging

1. INTRODUCTION

Statistical methods widely known under the name *kriging* are intended to predict unobserved values of a variable in a spatial domain, on the basis of observed values. These techniques are based on a function which describes the spatial dependence, the so-called *variogram*. Therefore, variogram estimation and variogram fitting are important stages of spatial prediction. Because they determine the kriging weights,

* Corresponding author. E-mail: reinhard.furrer@epfl.ch, genton@math.mil.edu.

they must be carried out carefully, otherwise kriging can produce unreliable maps.

In practical situations, a fraction of outliers is often included in observed data. Experience from a broad spectrum of applied sciences shows that measured data contain as a rule between 10 and 15 percent of outlying values due to gross errors, measurement mistakes, faulty recordings, etc. One might argue that any reasonable exploratory data analysis would identify and remove outliers in the data. However, this approach is often subjective and outlier rejection is highly opinion dependent. Thus, in this paper, we advocate the use of robust geostatistical methods, which prevent the negative effects of outlying values. Note that the existence of exploratory techniques does not supersede the utility of robust techniques.

The data set contains $n = 293$ measurements of the eight metals Zn, Cr, Cd, Co, Sr, Hg, Pb and Ni taken from Lake Geneva sediments. It has been provided by the C.I.P.E.L. (Commission Internationale de Protection des Eaux du Léman), an international commission on water quality control in Lake Geneva. Figure 1 depicts a map of the area of

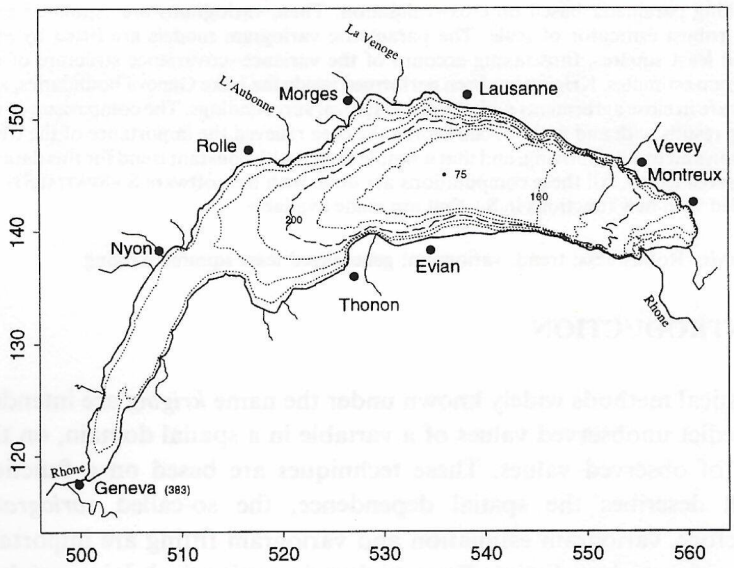

FIGURE 1 Map of the area of Lake Geneva in Switzerland. The main cities and rivers are represented, as well as elevation contour lines of the lake. The Rhone (main river) enters in the east part of the lake and leaves in the west part.

Lake Geneva in Switzerland. The main cities and rivers are represented, as well as elevation contour lines of the lake. Lausanne is the biggest city on the edge of the lake and possesses a huge purification plant located west of Lausanne on the border of the lake. Thus, sediments in this region often contain high values of metals. The Rhone is the main river of Lake Geneva. It enters in the east part of the lake and leaves in the west part. As several chemical companies are located along its course, high values of metals are also observed near its orifice area. Figure 2 shows the locations of the different measurements within a basic map of the area of Lake Geneva. The sampling is nearly regular and eight variables have been measured at each location point. Because the measured data have different units, we centered and reduced the observations (Mardia *et al.*, 1979) for the ongoing work.

In order to describe the data, the following model is used for each variable:

$$Z(\mathbf{x}) = m(\mathbf{x}) + \epsilon(\mathbf{x}), \quad \mathbf{x} = (x, y)^{\mathrm{T}}, \tag{1}$$

where $m(\mathbf{x})$ is the deterministic part of Z and $\epsilon(\mathbf{x})$ the stochastic one. Because of the local behavior of the data, the trend $m(\mathbf{x})$ is computed

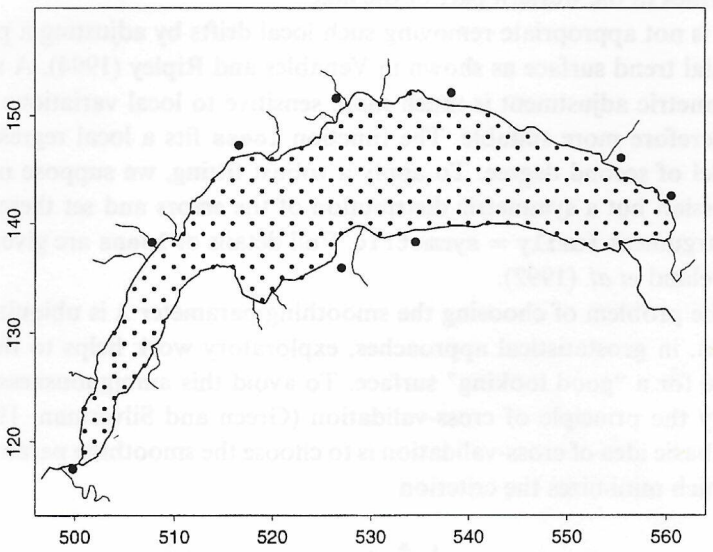

FIGURE 2 The $n=293$ locations of the data set within the lake boundaries. The sampling is nearly regular. Eight variables have been measured at each location point.

by a nonparametric approach and removed, as shown in the next section. Highly robust variogram estimation is performed in Section 3, followed by generalized least squares variogram fitting in Section 4. Finally, kriging results are discussed in the last part of this paper.

2. TREND DETECTION

The first step of spatial data analysis consists in detecting the trend of the variables, i.e. we determine $m(\mathbf{x})$, the nonstochastic part of (1), for each variable (Cressie, 1991). Figure 3 shows the trend surfaces of the eight variables of the data set drawn with the command `symbols` in S+. We note that each metal has a different global trend, but locally there are several similarities. By comparing the trends with the geographic and geologic characteristics of the neighborhood of the lake in Fig. 1, we verify that several variables (for example Zn, Cd, Co or Hg) have higher densities close to river estuaries and close to urban areas like Lausanne or Morges. The variables Zn, Cd and Pb are significantly higher in the neighborhood of the purification plant of Lausanne. Note also the contrasting spatial pattern of the variable Sr, probably due to industrial activities in the western part of the lake.

It is not appropriate removing such local drifts by adjusting a polynomial trend surface as shown in Venables and Ripley (1994). A nonparametric adjustment is much more sensitive to local variations and is therefore more suitable. The function `loess` fits a local regression model of second degree. To apply a robust fitting, we suppose not a Gaussian but a symmetric distribution of the errors and set therefore the argument `family = symmetric`. Full details of `loess` are given by Cleveland *et al.* (1992).

The problem of choosing the smoothing parameter λ is ubiquitous. Often, in geostatistical approaches, exploratory work helps to find a value for a "good looking" surface. To avoid this ambiguousness, we apply the principle of cross-validation (Green and Silverman, 1994). The basic idea of cross-validation is to choose the smoothing parameter λ which minimizes the criterion

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i - \hat{g}^{(-i)}(\mathbf{x}_i, \lambda)\right)^2, \tag{2}$$

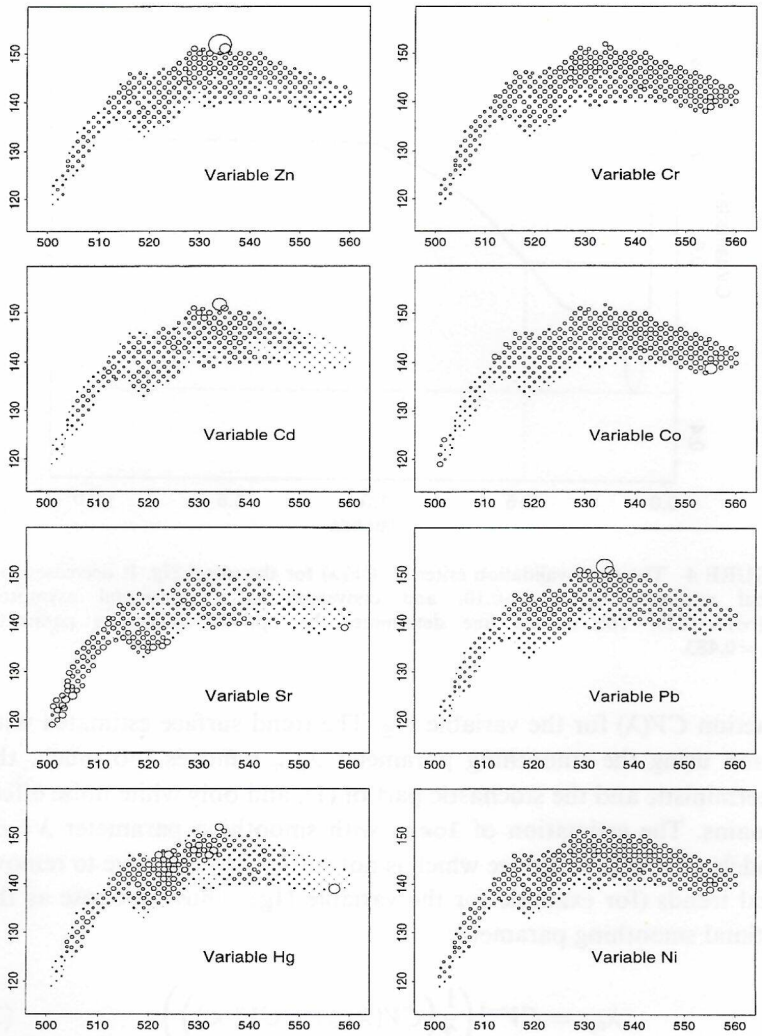

FIGURE 3 Trends for the eight variables of the data set. Bigger circles represent higher density.

where $\hat{g}^{(-i)}(\mathbf{x}_i, \lambda)$ is the estimation of $m(\mathbf{x}_i)$ in omitting the observation $\mathbf{x}_i$ from the data set. This criterion ensures stability of the fitted surface. In this work, the predictor $\hat{g}(\mathbf{x}, \lambda)$ is the function `loess`. In general, the function $CV(\lambda)$ decreases to a global minimum at $\lambda_{\min}$ close to 0, and converges to a horizontal asymptote $CV(\infty)$. Figure 4 shows the

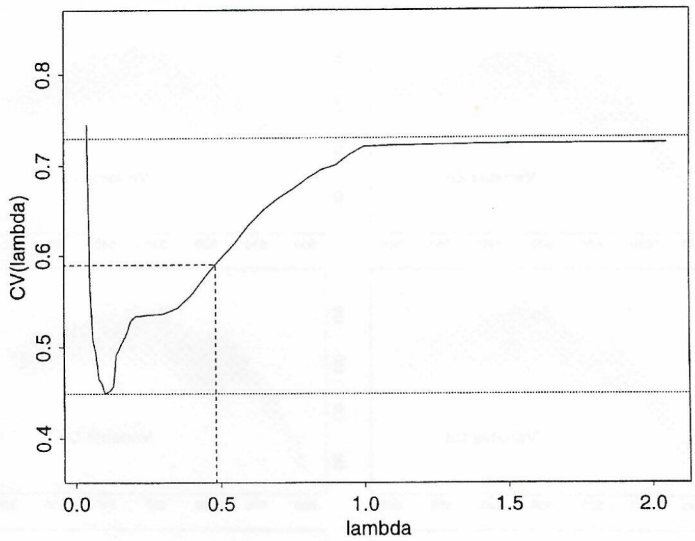

FIGURE 4 The cross-validation criterion $CV(\lambda)$ for the metal Hg. It decreases to a global minimum at $\lambda_{\min}=0.10$, and converges to a horizontal asymptote $CV(\infty)=0.729$. The dashed line determines the optimal smoothing parameter $\lambda_{\text{opt}}=0.483$.

function $CV(\lambda)$ for the variable Hg. The trend surface estimated with `loess` using the smoothing parameter $\lambda_{\min}$ removes too much, the deterministic and the stochastic part of (1), and only white noise effect remains. The estimation of `loess` with smoothing parameter $\lambda=\infty$ models a parabolic surface which is not sufficiently effective to remove local trends (for example for the variable Hg). Thus we chose as the optimal smoothing parameter

$$\lambda_{\text{opt}} = CV^{-1}\left(\frac{1}{2}\left(CV(\lambda_{\min}) + CV(\infty)\right)\right), \tag{3}$$

which is an intuitive compromise between irregularity and smoothness of the trend. Table I shows the values of $\lambda_{\min}$, $CV(\lambda_{\min})$, λ_{opt}, $CV(\lambda_{\text{opt}})$ calculated for the eight variables. Thus we removed the nonparametric trend $\hat{m}(\mathbf{x}) = \hat{g}(\mathbf{x}, \lambda_{\text{opt}})$ from each variable $Z(\mathbf{x})$ of the centered reduced data set. Note that the function `loess` usually fits a nonparametric model to independent and identically distributed (i.i.d.) observations. However, Lake Geneva sediments are typically dependent. Therefore,

TABLE I The values of the parameters $\lambda_{\min}$, $CV(\lambda_{\min})$, λ_{opt}, $CV(\lambda_{\text{opt}})$. The optimal smoothing parameter λ_{opt} is given in bold

Variable	$\lambda_{\min}$	$CV(\lambda_{\min})$	λ_{opt}	$CV(\lambda_{\text{opt}})$
Zn	0.059	0.276	**0.064**	0.469
Cr	0.110	0.335	**0.701**	0.390
Cd	0.083	0.427	**0.343**	0.511
Co	0.112	0.269	**0.745**	0.364
Sr	0.103	0.371	**0.200**	0.431
Pb	0.078	0.305	**0.091**	0.431
Hg	0.100	0.449	**0.483**	0.589
Ni	0.112	0.373	**0.854**	0.475

we introduced an optimal smoothing parameter λ_{opt} in order to fit a nonparametric trend to dependent observations.

3. HIGHLY ROBUST VARIOGRAM ESTIMATION

Variogram estimation is a crucial stage of spatial prediction, because it determines the kriging weights. It is important to have a variogram estimator which remains close to the true underlying variogram, even if outliers (faulty observations) are present in the data. Otherwise kriging can produce noninformative maps. Let $\epsilon(\mathbf{x}) = Z(\mathbf{x}) - m(\mathbf{x})$ be the detrended spatial stochastic process, which is assumed to be intrinsically stationary. The classical variogram estimator of a sample $\epsilon(\mathbf{x}_1), \ldots, \epsilon(\mathbf{x}_n)$ proposed by Matheron (1962), based on the method-of-moments, is

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N_{\mathbf{h}}} \sum_{N(\mathbf{h})} (\epsilon(\mathbf{x}_i) - \epsilon(\mathbf{x}_j))^2, \quad \mathbf{h} \in \mathbb{R}^d, \tag{4}$$

where $N(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j): \mathbf{x}_i - \mathbf{x}_j = \mathbf{h}\}$ and $N_{\mathbf{h}}$ is the cardinality of $N(\mathbf{h})$. This estimator is unbiased, but behaves poorly if there are outliers in the data. One single outlier can destroy this estimator completely. However, it is not enough to make simple modifications to formula (4), such as the ones proposed by Cressie and Hawkins (1980), in order to achieve robustness. In this section, we advocate the use of a highly robust variogram estimator (Genton, 1996; 1998a)

$$2\hat{\gamma}(\mathbf{h}) = (Q_{N_{\mathbf{h}}})^2, \quad \mathbf{h} \in \mathbb{R}^d, \tag{5}$$

which takes account of all the available information in the data. It is based on the sample $V_1(\mathbf{h}), \ldots, V_{N_{\mathbf{h}}}(\mathbf{h})$ from the process of differences

$V(\mathbf{h}) = \epsilon(\mathbf{x}+\mathbf{h}) - \epsilon(\mathbf{x})$ and the robust scale estimator $Q_{N_\mathbf{h}}$, proposed by Rousseeuw and Croux (1992; 1993):

$$Q_{N_\mathbf{h}} = 2.2191\{|V_i(\mathbf{h}) - V_j(\mathbf{h})|; i<j\}_{(k)}, \tag{6}$$

where the factor 2.2191 is for consistency at the Gaussian distribution, $k = \binom{[N_\mathbf{h}/2]+1}{2}$, and $[N_\mathbf{h}/2]$ denotes the integer part of $N_\mathbf{h}/2$. This means that we sort the set of all absolute differences $|V_i(\mathbf{h}) - V_j(\mathbf{h})|$ for $i<j$ and then compute its kth quantile ($k \approx 1/4$ for large $N_\mathbf{h}$). This value is multiplied by the factor 2.2191, thus yielding $Q_{N_\mathbf{h}}$. Note that this estimator computes the kth order statistic of the $\binom{N_\mathbf{h}}{2}$ interpoint distances. At first sight, the estimator $Q_{N_\mathbf{h}}$ appears to need $O(N_\mathbf{h}^2)$ computation time, which would be a disadvantage. However, it can be computed using no more than $O(N_\mathbf{h}\log N_\mathbf{h})$ time and $O(N_\mathbf{h})$ storage, by means of the fast algorithm described in Croux and Rousseeuw (1992).

This variogram estimator possesses several interesting properties of robustness. For instance, its influence function, which describes the effect on the estimator of an infinitesimal contamination, is bounded. This means that the worst influence that a small amount of contamination can have on the value of the estimator is finite, in opposition to Matheron's classical variogram estimator. Another important robustness property is the breakdown point ϵ^* of a variogram estimator, which indicates, how many data points need to be replaced to make the estimator explode (tend to infinity) or implode (tend to zero). The highly robust variogram estimator has an $\epsilon^* = 50\%$ breakdown point on the differences $V(\mathbf{h})$, the highest possible value, whereas Matheron's classical variogram estimator has only an $\epsilon^* = 0\%$ breakdown point, the lowest possible value. More details about the use and properties of this estimator, including some simulation studies, are presented in Genton (1998a).

As the data locations in east–west (E–W) and north–south (N–S) directions are spaced by two units, we estimate the variogram at even lags, with the common tolerance of one unit to achieve higher robustness. We compute the directional variograms for N–S and the E–W direction, as well as the omnidirectional variogram with (5) by using `variogram.qn`, a new function in S+. Note that S+SpatialStats uses originally only Matheron's classical variogram estimator (4) or Cressie and Hawkins' one. To compare the different variograms, Fig. 5 shows

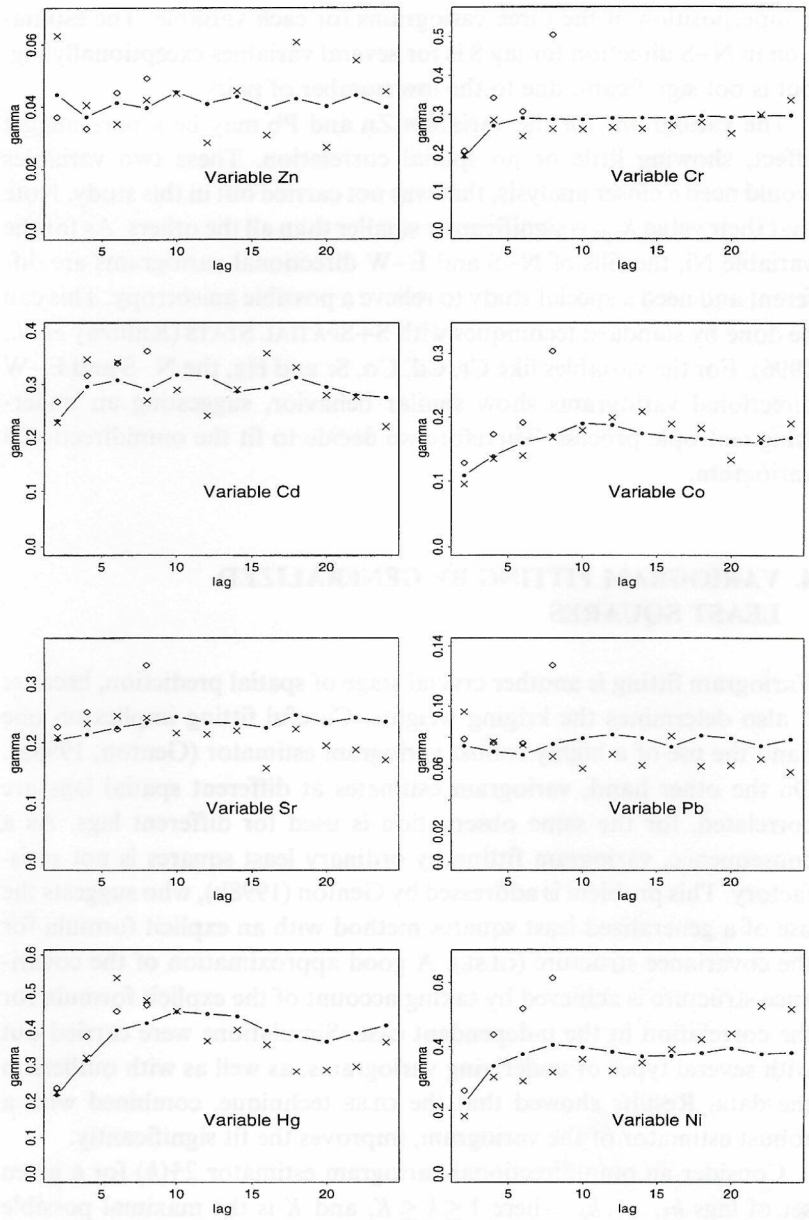

FIGURE 5 Empirical variograms for the eight variables. The north–south (N–S) directional variogram is marked with diamonds, the east–west (E–W) with crosses. The omnidirectional variogram is represented by jointed dots.

a superposition of the three variograms for each variable. The estimation in N–S direction for lag 8 is for several variables exceptionally big, but is not significant, due to the low number of pairs.

The variograms for the variables Zn and Pb may be a pure nugget effect, showing little or no spatial correlation. These two variables would need a closer analysis, that was not carried out in this study. Note that their value λ_{opt} is significantly smaller than all the others. As for the variable Ni, the sills of N–S and E–W directional variograms are different and need a special study to relieve a possible anisotropy. This can be done by standard techniques with S+SPATIAL STATS (Kaluzny *et al.*, 1996). For the variables like Cr, Cd, Co, Sr and Hg, the N–S and E–W directional variograms show similar behavior, suggesting an underlying isotropic process. Therefore we decide to fit the omnidirectional variogram.

4. VARIOGRAM FITTING BY GENERALIZED LEAST SQUARES

Variogram fitting is another crucial stage of spatial prediction, because it also determines the kriging weights. Careful fitting implies on one hand the use of a highly robust variogram estimator (Genton, 1998a). On the other hand, variogram estimates at different spatial lags are correlated, for the same observation is used for different lags. As a consequence, variogram fitting by ordinary least squares is not satisfactory. This problem is addressed by Genton (1998b), who suggests the use of a generalized least squares method with an explicit formula for the covariance structure (GLSE). A good approximation of the covariance structure is achieved by taking account of the explicit formula for the correlation in the independent case. Simulations were carried out with several types of underlying variograms, as well as with outliers in the data. Results showed that the GLSE technique, combined with a robust estimator of the variogram, improves the fit significantly.

Consider an omnidirectional variogram estimator $2\hat{\gamma}(h)$ for a given set of lags $h_1, \ldots, h_k$, where $1 \leq k \leq K$, and K is the maximal possible distance between data. Denote further by $2\hat{\boldsymbol{\gamma}} = (2\hat{\gamma}(h_1), \ldots, 2\hat{\gamma}(h_k))^{\mathrm{T}} \in \mathbb{R}^k$ the random vector with the variance–covariance matrix $\mathrm{Var}(2\hat{\boldsymbol{\gamma}}) = \tau^2\Omega$, where τ^2 is a real positive constant. Suppose that one

wants to fit a valid parametric variogram $2\gamma(h, \boldsymbol{\theta})$ to the estimated points $2\hat{\boldsymbol{\gamma}}$. The method of generalized least squares consists in determining the estimator $\hat{\boldsymbol{\theta}}$ which minimizes

$$G(\boldsymbol{\theta}) = (2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\theta}))^{\mathrm{T}} \Omega^{-1} (2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\theta})), \tag{7}$$

where $2\boldsymbol{\gamma}(\boldsymbol{\theta}) = (2\gamma(h_1, \boldsymbol{\theta}), \ldots, 2\gamma(h_k, \boldsymbol{\theta}))^{\mathrm{T}} \in \mathbb{R}^k$ is the vector of the valid parametric variogram, and $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter to be estimated. Note that $2\gamma(h, \boldsymbol{\theta})$ is generally a nonlinear function of the parameter $\boldsymbol{\theta}$. Journel and Huijbregts (1978) suggest to use only lag vectors h_i such that $N_{h_i} > 30$ and $0 < i \leq K/2$. This empirical rule is often met in practice, and is also used in this work. The GLSE algorithm is the following:

Step 1: Determine the matrix $\Omega = \Omega(\boldsymbol{\theta})$ with element Ω_{ij} given by

$$\mathrm{Corr}(2\hat{\gamma}(h_i), 2\hat{\gamma}(h_j))\gamma(h_i, \boldsymbol{\theta})\gamma(h_j, \boldsymbol{\theta}) \Big/ \sqrt{N_{h_i} N_{h_j}}. \tag{8}$$

Step 2: Choose $\boldsymbol{\theta}^{(0)}$ and let $l = 0$.

Step 3: Compute the matrix $\Omega(\boldsymbol{\theta}^{(l)})$ and determine $\boldsymbol{\theta}^{(l+1)}$ which minimizes

$$G(\boldsymbol{\theta}) = (2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\theta}))^{\mathrm{T}} \Omega(\boldsymbol{\theta}^{(l)})^{-1} (2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\theta})).$$

Step 4: Repeat Step 3 until convergence to obtain $\hat{\boldsymbol{\theta}}$.

In Step 1, the correlation $\mathrm{Corr}(2\hat{\gamma}(h_i), 2\hat{\gamma}(h_j))$ can be approximated by the one in the independent case. An explicit formula can be found in Genton (1998b), which depends only on the lags h_i and h_j, as well as on the size $n = n_1 n_2$ of a spatial rectangular data set. In Step 2, the choice of $\boldsymbol{\theta}^{(0)}$ can be carried out randomly, or with the result of a fit by ordinary least squares (OLS).

A spherical variogram

$$\gamma(h, \boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0, \\ \theta_1 + \theta_2 \left(\frac{3}{2}(h/\theta_3) - \frac{1}{2}(h/\theta_3)^3 \right) & \text{if } 0 < h \leq \theta_3, \\ \theta_1 + \theta_2 & \text{if } h > \theta_3, \end{cases}$$

TABLE II The estimated parameters $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ by GLSE of the spherical variogram

Variable	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
Zn	0.038	0.003	7.076
Cr	0.078	0.213	5.669
Cd	0.119	0.177	3.948
Co	0.074	0.096	7.189
Sr	0.189	0.047	10.384
Pb	0.069	0.009	12.665
Hg	0.074	0.328	6.513
Ni	0.071	0.312	5.802

has been fitted to the empirical omnidirectional variograms by GLSE using `glse.fitting`, another new function in S+. The starting value $\boldsymbol{\theta}^{(0)}$ was set as the solution of a fit by OLS. The estimated parameters $\hat{\boldsymbol{\theta}}$ of the fitting variograms are given in Table II. To calculate (8), we neglected the irregularities of the grid and set $n_1 = 9$ and $n_2 = 32$, which is a reasonable approximation of the grid.

5. KRIGING

Epitomizing, kriging is a linear interpolation method that allows predictions of unknown values of a random function from observations at known locations. For further details see Cressie (1991). S+SPATIALSTATS performs two-dimensional kriging by using the `krige` and `predict.krige` functions. The kriging results are easily visualized with the functions `contour` or `persp`. Further details and examples are given by Kaluzny *et al.* (1996). Figures 6 and 7 shows the kriging maps for the centered reduced variables Cr, Co, Sr and Hg. The kriging predictions are calculated over a 100×100 grid inside the lake boundaries.

In order to compare the model (1) with a nonparametric trend $m(\mathbf{x})$ to another one with a constant trend $m(\mathbf{x}) = m$, and thus to relieve the importance of trend removing for this data set, we compare the kriging results of the variable Cd without and with trend removing. Figure 8 shows the empirical and the fitted variograms of the variable Cd on the centered reduced data and on the data with removed trend. The difference on the estimated parameter $\hat{\boldsymbol{\theta}}$ is remarkable, especially on the

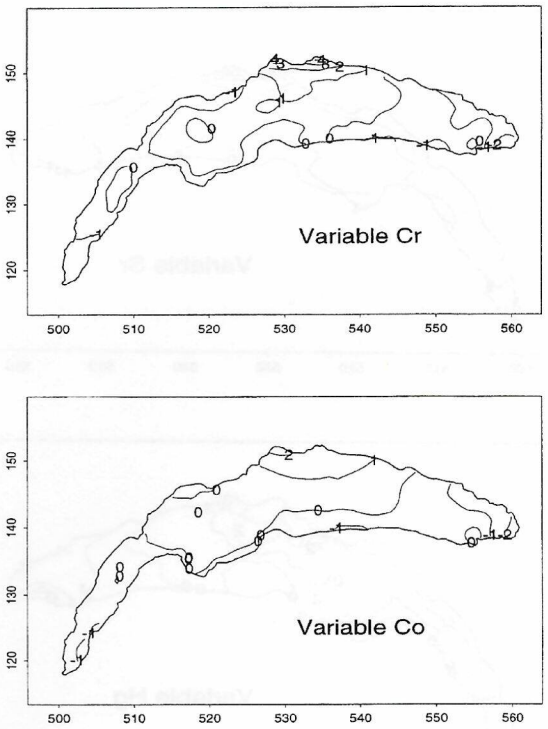

FIGURE 6 Kriging maps for the centered reduced variables Cr and Co. Kriging predictions are calculated over a 100×100 grid inside the lake boundaries. The contour lines are plotted for the levels $-2, -1, \ldots, 4$.

range θ_3 of the variogram: $\hat{\boldsymbol{\theta}} = (0.151, 0.936, 27.407)$ on the centered reduced data set versus $\hat{\boldsymbol{\theta}} = (0.119, 0.177, 3.948)$ on the data set with removed trend. These two parameters have been used to predict the values over a 100×100 grid inside the lake boundaries. Figure 9 shows the kriging results, which are locally significantly different. Effectively, even though the two kriging maps are not completely unalike (due to the exact interpolation of the kriging estimator, as well as the high number of regular sampling locations), the differences are more noticeable at a smaller scale. As the kriging is an exact predicting method, situations with a small number of observations are more affected by differences between variograms.

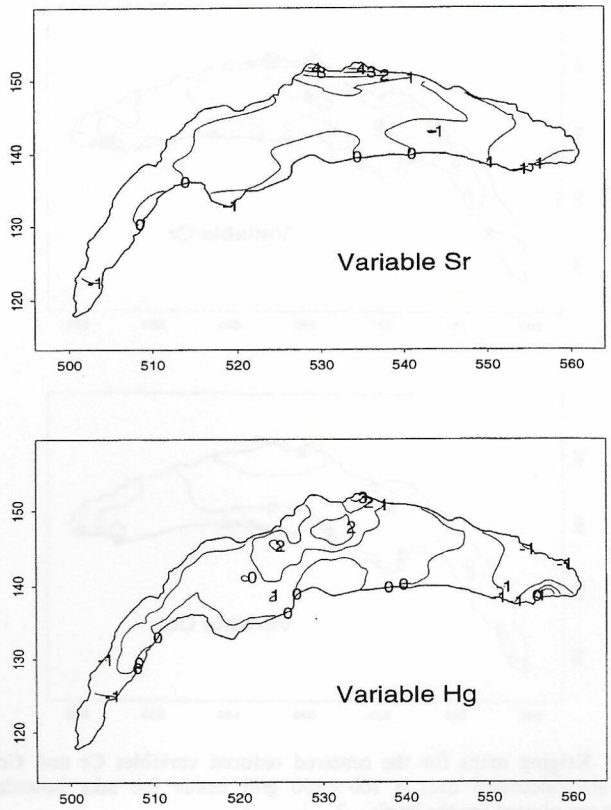

FIGURE 7 Kriging maps for the centered reduced variables Sr and Hg. Kriging predictions are calculated over a 100×100 grid inside the lake boundaries. The contour lines are plotted for the levels $-1, 0, \ldots, 4$.

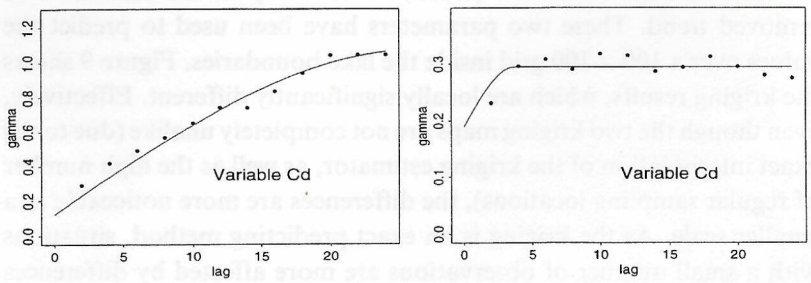

FIGURE 8 Empirical and fitted variograms of the variable Cd on centered reduced data (left side) and on data with removed trend (right side).

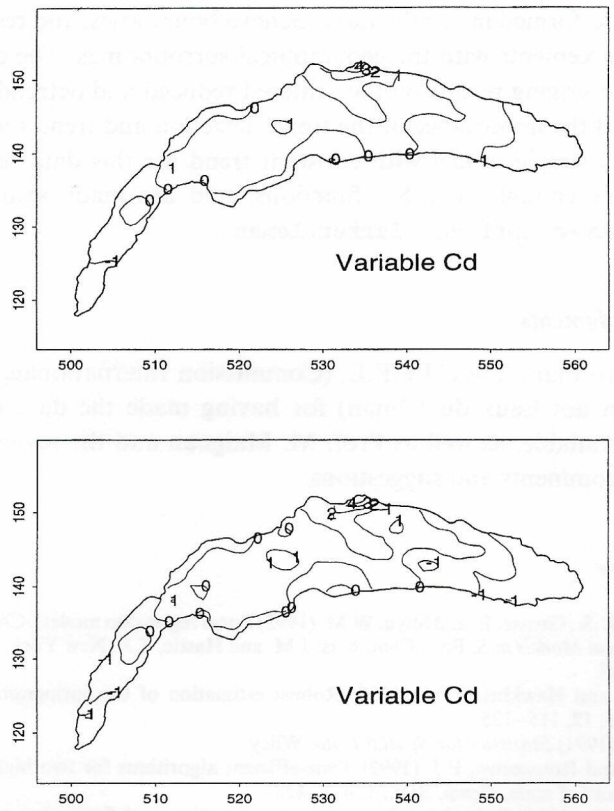

FIGURE 9 Kriging map of the variable Cd on centered reduced data (top) and on data with removed trend (bottom). Kriging predicitions are calculated over a 100×100 grid inside the lake boundaries. The contour lines are plotted for the levels $-1, 0, \ldots, 4$.

6. CONCLUSIONS

In this paper we have studied a data set of sediments from Lake Geneva in Switzerland, which consists of 293 measurements of eight metals. As local drifts are typically present in the data set due to river estuaries and urban areas, each variable has been detrended by a nonparametric surface, based on a cross-validation criterion. Then, robust methods have been applied for variography with the software S+SPATIALSTATS. First, the variogram was estimated by a highly robust estimator. Second, the fit of the variogram estimates was done by generalized least squares thus taking account of their statistical properties. Kriging

has been performed inside the Lake Geneva boundaries, and results are in close agreements with the geographical surroundings. The comparison of the kriging results on the centered reduced and detrended data set relieved the importance of the trend detection and trend removing, and that a simple model with constant trend for this data set is not satisfactory enough. The S+ functions used are made available at `http://dmawww.epfl.ch/~furrer/leman/`.

Acknowledgments

We wish to thank the C.I.P.E.L. (Commission Internationale pour la Protection des Eaux du Léman) for having made the data of Lake Geneva available, as well as Prof. M. Maignan and the reviewers for valuable comments and suggestions.

References

Cleveland, W.S., Grosse, E. and Shyu, W.M. (1992) Local regression models. Chapter 8 in: *Statistical Models in S*. Eds., Chambers, J.M. and Hastie, T.J., New York: Chapman and Hall.

Cressie, N. and Hawkins, D.M. (1980) Robust estimation of the variogram I, *Math. Geology*, **12**, 115–125.

Cressie, N. (1991) *Statistics for Spatial Data*, Wiley.

Croux, C. and Rousseeuw, P.J. (1992) Time-efficient algorithms for two highly robust estimators of scale, *Comp. Stat.*, **1**, 411–428.

Genton, M.G. (1996) Robustness in variogram estimation and fitting in geostatistics, Ph.D. thesis #1595, Department of Mathematics, Swiss Federal Institute of Technology.

Genton, M.G. (1998a) Highly robust variogram estimation, *Math. Geology*, **30**, 213–221.

Genton, M.G. (1998b) Variogram fitting by generalized least squares using an explicit formula for the covariance structure, *Math. Geology*, **30**, 853–871.

Green, P.J. and Silverman, B.W. (1994) *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall.

Kaluzny, S.P., Vega, S.C., Cardoso, T.P. and Shelly, A.A. (1996) S+SpatialStats, User's Manual, MathSoft.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate Analysis*, Academic Press.

Matheron, G. (1962) Traité de géostatistique appliquée, Tome I. *Mémoires du Bureau de Recherches Géologiques et Minières*, **30**, Editions Technip, Paris.

Rousseeuw, P.J. and Croux, C. (1992) Explicit scale estimators with high breakdown point, L_1 *Statistical Analysis*, 77–92.

Rousseeuw, P.J. and Croux, C. (1993) Alternatives of the median absolute deviation, *Journal of the American Statistical Association*, **88**, 1273–1283.

Venables, W.N. and Ripley, B.D. (1994) *Modern Applied Statistics with S-Plus*, Springer.